

HETEROGENEITY AND AGGREGATION IN
SEASONAL TIME SERIES

By
Georgios Tripodis

SUBMITTED IN FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
AT
UNIVERSITY OF LONDON
LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE
DEPARTMENT OF STATISTICS
NOVEMBER 2006

UMI Number: U615884

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U615884

Published by ProQuest LLC 2014. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

THESES

F8743

Bib. 1125308



Abstract

Seasonality is an important part of many real time series. While issues of seasonal heteroscedasticity and aggregation have been a cause of concern for data users, there has not been a great deal of theoretical research in this area. This thesis concentrates on these two issues.

We consider seasonal time series with single season heteroscedasticity. We show that when only one month has different variability from others there are constraints on the seasonal models that can be used. We show that both the dummy and the trigonometric models are not effective in modelling seasonal series with this type of variability. We suggest two models that permit single season heteroscedasticity as a special case. We show that seasonal heteroscedasticity gives rise to periodic autocorrelation function. We propose a new class, called periodic structural time series models (PSTSM) to deal with such periodicities. We show that PSTSM have correlation structure equivalent to that of a periodic integrated moving average (PIMA) process. In a comparison of forecast performance for a set of quarterly macroeconomic series, PSTSM outperform periodic autoregressive (PAR) models both within and out of sample.

We also consider the problem of contemporaneous aggregation of time series using the structural time series framework. We consider the conditions of identifiability for the aggregate series. We show that the identifiability of the models for the component series is not sufficient for the identifiability of the model for the aggregate series. We also consider the case where there is no estimation error as well as the case of modeling an unknown process. For the case of the unknown process we provide recursions based on the Kalman filter that give the asymptotic variance of the estimated parameters.

Contents

1	Introduction and summary	14
2	Basic Concepts	19
2.1	Introduction	19
2.2	State space methods	20
2.3	Structural time series models	21
2.3.1	Dummy seasonality model (DS)	23
2.3.2	Trigonometric seasonal model (TS)	25
2.3.3	Harrison and Stevens seasonal model (HS)	28
2.4	Kalman filter and smoother	30
2.4.1	Derivation of Kalman filter	32
2.4.2	Fixed interval smoother	35
2.4.3	Disturbance smoother	39
2.5	Square root filter	43

2.5.1	Givens rotations	44
2.5.2	Square root form	45
2.6	Initialisation	46
2.7	Estimation	48
2.7.1	Numerical optimisation algorithms	48
2.7.2	The score vector	50
2.8	Diagnostic checking and model selection	52
3	Periodic variance in one season	55
3.1	Introduction	55
3.2	Models for single season heteroscedasticity	57
3.2.1	Periodic variance in the seasonal component	58
3.2.2	A comparison of periodic seasonal and periodic irregular models	60
3.2.3	Other seasonal models	64
3.2.4	Score vectors for seasonal models	68
3.2.5	Initialisation of HS	72
3.3	Test for seasonal heteroscedasticity	72
3.4	Applications	78
3.5	Conclusions	89
4	Structural time series models for periodic processes	91

4.1	Introduction	91
4.2	Periodic AR and MA models	93
4.3	Periodic structural times series models	96
4.3.1	Periodic local level model	96
4.3.2	Periodic local linear trend model	100
4.3.3	Periodic basic structural model	102
4.4	Applications	105
4.5	Conclusions	112
5	Contemporaneous aggregation of time series	113
5.1	Introduction	113
5.2	Aggregation of time series	115
5.3	Comparison of forecasts	125
5.4	Estimated coefficients	127
5.5	Application	131
5.6	Conclusions	133
6	Conclusion and further research	136
A		140
A.1	Givens Rotations algorithm	140
A.2	Square Kalman Filter code	143

List of Figures

3.1	Graphic diagnostics for seasonal heteroscedasticity (index of production for Italy)	74
3.2	Time series used in applications section : index of production for Italy, France and Spain, and Whiterocks river flow	84
3.3	Residuals diagnostics for index of production for Italy	85
3.4	Residuals diagnostics for index of production for France	86
3.5	Residuals diagnostics for index of production for Spain	87
3.6	Residuals diagnostics for Whiterocks river flow	88
4.1	Time series used in the application of periodic models (1)	107
4.2	Time series used in the application of periodic models (2)	108
5.1	Motor Vehicle Production in the UK	134

List of Tables

3.1	Lags with periodic autocorrelation in stationary form	64
3.2	Rejection frequency (%) for periodic irregular variance model .	75
3.3	Rejection frequency (%) for periodic seasonal variance model .	76
3.4	Rejection frequency (%) for periodic irregular variance model : comparing across signal to noise ratios	76
3.5	Rejection frequency (%) for periodic seasonal variance model : comparing across signal to noise ratios	77
3.6	Italian Index of Production: parameter estimates and diag- nostic statistics	80
3.7	French Index of Production: parameter estimates and diag- nostic statistics	81
3.8	Spanish Index of Production: parameter estimates and diag- nostic statistics	82
3.9	River flow in Whiterocks river: parameter estimates and di- agnostic statistics	83
4.1	Comparison of PAR and PSTSM (log is applied to all the series except CAN Unemployment)	110

4.2	Out of Sample Comparison of PAR and STSM (RMSE)	111
5.1	Comparison of direct and indirect methods	133

To my Grandfather

Ithaca

When you set out on your journey to Ithaca,
pray that the road is long,
full of adventure, full of knowledge.
The Lestrygonians and the Cyclops,
the angry Poseidon – do not fear them:
You will never find such as these on your path,
if your thoughts remain lofty, if a fine
emotion touches your spirit and your body.
The Lestrygonians and the Cyclops,
the fierce Poseidon you will never encounter,
if you do not carry them within your soul,
if your soul does not set them up before you.

Pray that the road is long.
That the summer mornings are many, when,
with such pleasure, with such joy you will enter ports seen for the first time;
stop at Phoenician markets,
and purchase fine merchandise,
mother-of-pearl and coral, amber and ebony,
and sensual perfumes of all kinds,
as many sensual perfumes as you can;
visit many Egyptian cities,
to learn and learn from scholars.

Always keep Ithaca in your mind.

To arrive there is your ultimate goal.
But do not hurry the voyage at all.
It is better to let it last for many years;
and to anchor at the island when you are old,
rich with all you have gained on the way,
not expecting that Ithaca will offer you riches.

Ithaca has given you the beautiful voyage.
Without her you would have never set out on the road.
She has nothing more to give you.

And if you find her poor, Ithaca has not deceived you.
Wise as you have become, with so much experience,
you must already have understood what Ithacas mean.

Constantine P. Cavafy (1911)

Acknowledgements

I came to Britain nine years ago. During this time I learned many things about life. I loved some people, hurt few others and hated a couple. Overall it was the most wonderful stop in my journey up to now. While I was here, I learned also a few things about Time Series, some of which I tried to write up in this thesis. It is customary the acknowledgments section of a PhD thesis to resemble an Oscar acceptance speech. I cannot escape from this honourable tradition. It is hard to write down all the people that made this journey enjoyable. First of all I would like to thank my supervisor Jeremy Penzer for his patience and support while trying to teach me some Time Series. I owe a lot to Simon Compton for, among other things, showing me how to analyse time series in practice. I would like to thank Professor Wynn and Dr Finkenstadt for their kind remarks and helpful comments while acting as examiners for this thesis.

I cannot forget to mention Katerina, Cecile and Michelle for their love and support in different periods of my life. There is a huge list of friends that I would like to thank. So thanks to Panos, Diego and Jaya: we had the most interesting conversations in the last 4 years, on Statistics, Life and the Secret of the Universe (I am not going to reveal it, otherwise it will stop being a secret). It would be unfair not to mention Adrian because he is *einfach unbeschreiblich*. I would like also to thank Anca, Claudia and Francis from ONS because I am lucky to have them as friends. Special thanks to Anita (because she told me so, and I have learned to do exactly what she says).

Last but not least, my Annushka for the unlimited love she has shown. From the depths of my heart I would like to express my gratitude to the medical staff of Kathleen Ferrier Ward and the Urology department of the Middlesex hospital. It might sound banal, but in this occasion it is true that without them this thesis would not have been finished.

Finally, thanks Mum!! You are the Greatest!! I dedicate this thesis to my Grandad. He taught me to read and write 28 years ago during a hot summer in Itea. I am sure he would have been proud to see that his efforts were not futile.

The Journey continues...

London, England
August 15, 2005

Yorghos Tripodis

Chapter 1

Introduction and summary

Seasonality is an important part of many real time series. From births and deaths to the Gross Domestic Product of any country, seasonal effects are prominent. Analysts attempt to understand and estimate seasonal effects in order to either remove them through seasonal adjustment or forecast them. In the last 10 years, there has been considerable progress in research on seasonal time series. This is partly a result of growing interest from national statistical institutes that publish most series in a seasonally adjusted form. The US Bureau of Census has developed X-12-ARIMA (Findley, Monsell, Otto, Bell, and Pugh 1998) based on the well known X-11 method. On the other hand, Eurostat has concentrated research on TRAMO-SEATS (Gomez and Maravall 1996), which is based on an ARIMA model decomposition. Recently the two methods have been brought together within the same platform called X-13A-S (Findley 2005). Both these lines of research concentrate on identifying a relatively stable seasonal component and finding the best way to remove it. While issues of seasonal heteroscedasticity and aggregation

have been a cause of concern for practitioners in these institutes as well as data users, there has not been a great deal of theoretical research in this area. This thesis concentrates on these two issues.

Seasonal heteroscedasticity is evident in many economic time series. The existence of seasonal heteroscedasticity can be attributed to a combination of economic behaviour and administrative practices. For example, the Average Earning Index for United Kingdom shows a higher volatility in the months of March and December. This is a result of the so-called “bonus effect”; the big financial corporations of the City of London give large bonuses to some of their employees at the end of the financial year and to a lesser extent at the end of the calendar year. This brings up the total index for the average earnings for the whole of UK in these two months. The size of the bonuses are directly linked with the state of the economy since a better economic outlook will bring higher profits and therefore higher bonuses. In this example, the seasonal component is linked with the business cycle, creating seasonal heteroscedasticity.

The second issue we consider is that of aggregated time series. This is an important issue in many practical applications of national statistics. For example, motor vehicle production index in the UK is broken down into production for home and export markets. Until recently, all three series have been seasonally adjusted separately, but following a methodological review (Tripodis 2005) the export series is derived from the total series and the home market series. The problem is related to whether it is better to forecast the components of a dataset and add up the forecasts (indirect method), or to

forecast the aggregate series separately (direct method).

Chapter 2 provides the reader with the necessary background on Structural Time Series Models (STSM). Popularised by Harvey (1989), this family of models provides clear advantages over the autoregressive integrated moving averages (ARIMA) class of models, particularly in the analysis of economic time series. We use STSM as the main tool for analysing time series. STSM allow us to model directly the salient features of a dataset as simple stochastic processes. This chapter introduces the state space form, in which structural time series models are generally written. Once written in a state space form, estimation of the structural models is straight-forward. The Kalman filter provides one-step ahead prediction errors along with their associated variances which can then be plugged into the prediction error decomposition of the likelihood function. The likelihood function is then maximised with any of the widely used maximisation methods, such as Newton's method.

Chapter 3 introduces the problem of seasonal heteroscedasticity. We begin by looking at the simple case where a single season has different volatility compared to all other seasons. This behaviour is a feature of many economic time series. For example in monthly production series, the variability is higher for the month with the lowest level of production. We show that modeling of this single seasonal heteroscedasticity is more complicated than originally thought and only few seasonal models can be used to model this behaviour. We look at the power of a likelihood ratio test for identifying single season heteroscedasticity. We also look at some real life seasonal time series, showing the relative merits of some seasonal models with seasonal

heteroscedasticity for different applications.

Chapter 4 generalises the concept of seasonal heteroscedasticity to many periods and to different components. We show that seasonal heteroscedasticity can be identified by periodicity in the autocorrelation function. We define the periodic structural time series models which can be used to model periodic autocorrelation. We show that the periodic structural time series models are equivalent to periodic moving average models. The advantage of the structural approach is the ease of identifying the appropriate model. We compare the periodic structural models with periodic autoregressive models which have been used extensively in economic time series (Franses 1996). We show that in most cases, the periodic structural models provide better forecasts than the periodic autoregressions for a set of macroeconomic time series.

Chapter 5 looks at the problem of contemporaneous aggregation of time series. Extensive work has been done in this area for ARIMA models. In this thesis, we consider the structural time series framework. We show different ways of aggregating time series models and consider the conditions of identifiability for the aggregate series. We show that the identifiability of the models for the component series is not sufficient for the identifiability of the model for the aggregate series. We also consider the case where there is no estimation error as well as the case of modeling an unknown process. For the case of the unknown process we provide recursions based on the Kalman filter that give the asymptotic variance of the estimated parameters.

Finally chapter 6 provides the conclusions and some ideas for further research. The Appendix includes some subroutines written in Ox (Doornik 1998) that were used in the estimation of the models described in this thesis.

Chapter 2

Basic Concepts

2.1 Introduction

In this chapter we provide theoretical background to the thesis. Our focus is on unobserved component models and in particular on structural time series models (STSMs). Popularized by Harvey (1989), STSM provides clear advantages over the autoregressive integrated moving averages (ARIMA) class of models, particularly in the analysis of economic time series. STSMs are readily identified and their parameters provide information about salient features such as trend and seasonality. In §2.3, the structural time series models used in the applications of this thesis are defined. The state space form is defined in §2.2 where some examples are also given. Since this thesis concentrates on the seasonal behaviour of a time series, we define and sketch in §2.3.1 to §2.3.3 the main characteristics of widely used seasonal models. The Kalman filter provides the basis for inference in any model that is cast in

state-space form. Filtering and smoothing algorithms are introduced in §2.4; filtering provides the best linear estimates of a system given the previous observations, while smoothing provides the best linear estimates of the system given the entire sample. Rounding errors and matrices close to singularity may cause the Kalman filter to break down. Under these circumstances a transformed version, known as the square root filter ensures that the state covariance matrix is always positive definite; details are given in §2.5 and in the appendix we provide a set of subroutines written in Ox (Doornik 1998) used in conjunction with Ssfpack (Koopman, Shephard, and Doornik 1998) implementing the square root filter. Ssfpack is a suite of C routines used for the statistical analysis of univariate and multivariate models which are cast in the state-space form. §2.6 discusses how the Kalman filter is initialised when the starting values for the state are unknown. §2.7 discusses how the structural models are estimated via the prediction error decomposition of the likelihood function. Parameter estimation requires numerical maximisation algorithms which are presented in §2.7.1. The final section presents the main diagnostic tools used for checking and model selection throughout this thesis.

2.2 State space methods

A structural time series model can be estimated once it is represented in state space form. Applying the Kalman filter and smoother to the state space form gives minimum mean square linear estimators of the components. Assume we have p time series, we denote by $y_{i,t}$ the observation of the i^{th} series at time t .

Then let $\mathbf{Y}_{i,t} = [y_{i,1}, \dots, y_{i,t}]'$, $\mathbf{y}_t = [y_{1,t}, \dots, y_{p,t}]'$ and $\mathbf{Y}_t = [\mathbf{Y}_{1,t}, \dots, \mathbf{Y}_{p,t}]$. A convenient representation of the linear state space model is (Durbin and Koopman 2001):

$$\begin{aligned} \mathbf{y}_t &= \mathbf{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\epsilon}_t && \text{observation equation} \\ \boldsymbol{\alpha}_{t+1} &= \mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{R}_t \boldsymbol{\eta}_t && \text{measurement equation} \end{aligned} \quad (2.2.1)$$

where $\boldsymbol{\epsilon}_t \sim \text{NID}(\mathbf{0}, \mathbf{H}_t)$, $\boldsymbol{\eta}_t \sim \text{NID}(\mathbf{0}, \mathbf{Q}_t)$, and NID denotes normally and independently distributed. Furthermore, \mathbf{y}_t is a $p \times 1$ vector of observations and $\boldsymbol{\alpha}_t$ is an unobserved $m \times 1$ vector called the state vector. The first equation can be seen as linear regression with time varying coefficients. The second equation assumes that the time-varying coefficients follow a Markov process. For the purpose of this thesis, we assume that $\{\boldsymbol{\epsilon}_t\}$, and $\{\boldsymbol{\eta}_t\}$ are uncorrelated. This assumption can be relaxed for general models. The matrices \mathbf{Z}_t , \mathbf{T}_t , \mathbf{Q}_t , \mathbf{R}_t , and \mathbf{H}_t are deterministic and depend on elements of an unknown parameter vector $\boldsymbol{\psi}$, estimated by maximum likelihood. The state space form can be used to represent a wide range of time series models.

2.3 Structural time series models

In an unobserved component model all components are modelled explicitly as stochastic processes. A key distinction for the structural time series model is that all components represent salient features of the data, such as trend. A detailed discussion of the structural time series models is found in Harvey (1989). In the structural model paradigm, a time series can be decomposed into its salient features such as trend, seasonal and business cycle component.

This decomposition can also be seen within the framework of factor analysis. Standard factor analysis, tries to determine m uncorrelated unobservable common factors which are linear combinations of the of n ($m < n$) observable correlated variables and explain the mutual correlation of the system. The aim of a univariate structural time series model is to determine m unobserved components of an observed time series $\{y_t : t = 1, \dots, n\}$ with correlated observations. The components are associated with the salient features of the time series and each observation of $\{y_t\}$ is the sum of the m unobserved components measured at time t . Each component at time t is a linear combination of future and past observations. In general we may assume that the components are mutually uncorrelated. Nevertheless, as in standard factor analysis, this can be extended to the case where some components are mutually correlated.

The simplest structural time series model is the local level model (LLM) in which the level of the series follows a random walk.

$$\begin{aligned} y_t &= \mu_t + \epsilon_t & \{\epsilon_t\} &\sim \text{NID}(0, \sigma_\epsilon^2) \\ \mu_{t+1} &= \mu_t + \eta_t & \{\eta_t\} &\sim \text{NID}(0, \sigma_\eta^2) \end{aligned}$$

where $\{\epsilon_t\}$, and $\{\eta_t\}$ are mutually uncorrelated. By adding a slope term $\{\beta_t\}$, which also follows a random walk, we obtain the local linear trend model:

$$\begin{aligned} y_t &= \mu_t + \epsilon_t & \{\epsilon_t\} &\sim \text{NID}(0, \sigma_\epsilon^2) \\ \mu_{t+1} &= \mu_t + \beta_t + \eta_t & \{\eta_t\} &\sim \text{NID}(0, \sigma_\eta^2) \\ \beta_{t+1} &= \beta_t + \zeta_t & \{\zeta_t\} &\sim \text{NID}(0, \sigma_\zeta^2) \end{aligned}$$

where $\{\mu_t\}$ is the trend. The matrices of the state-space form are:

$$\begin{aligned} \boldsymbol{\alpha}_t &= (\mu_t, \beta_t)' & \mathbf{Z}_t &= (1, 0) \\ \mathbf{H}_t &= \sigma_\epsilon^2 & \mathbf{R}_t &= \mathbf{I}_2 \end{aligned}$$

$$\mathbf{T}_t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

$$\mathbf{Q}_t = \begin{pmatrix} \sigma_\eta^2 & 0 \\ 0 & \sigma_\zeta^2 \end{pmatrix}$$

We are particularly interested in seasonal series. A standard decomposition that we will use in this chapter is:

$$y_t = \mu_t + \gamma_t + \epsilon_t \quad \{\epsilon_t\} \sim \text{NID}(0, \sigma_\epsilon^2) \quad (2.3.1)$$

where $\{\gamma_t\}$ is the seasonal component. The model given by (2.3.1) is referred to as the basic structural model (BSM). Descriptions of commonly used models for seasonality are given below.

2.3.1 Dummy seasonality model (DS)

A simple way to guarantee a deterministic seasonal pattern is to assume that the seasonal effects sum to zero, that is, $\gamma_t = -\sum_{j=1}^{s-1} \gamma_{t-j}$. We allow seasonality to evolve over time by adding a white noise term $\{\omega_t\}$. This gives the relationship (Harvey 1989):

$$\gamma_t = -\sum_{j=1}^{s-1} \gamma_{t-j} + \omega_t \quad \{\omega_t\} \sim \text{NID}(0, \sigma_\omega^2) \quad (2.3.2)$$

or equivalently:

$$S(L)\gamma_t = \omega_t$$

where $S(L)$ is the seasonal summation operator, $S(L) = \sum_{i=0}^{s-1} L^i$.

Subtracting γ_{t-s} from both sides of (2.3.2), we get:

$$\gamma_t - \gamma_{t-s} = - \sum_{j=1}^s \gamma_{t-j} + \omega_t \implies \Delta_s \gamma_t = \omega_t - \omega_{t-1} \quad (2.3.3)$$

since $S(L)\gamma_{t-1} = \omega_{t-1}$. Thus $\{\gamma_t\}$ follows a seasonal ARIMA $(0, 0, 1) \times (0, 1, 0)_s$.

For the case of a local level and a dummy seasonal component, we have:

$$y_t = \mu_t + \gamma_t + \epsilon_t \quad (2.3.4)$$

$$\mu_t = \mu_{t-1} + \eta_t$$

$\{\gamma_t\}$ follows (2.3.2), while $\{\epsilon_t\} \sim \text{NID}(\sigma_\epsilon^2)$, $\{\omega_t\} \sim \text{NID}(0, \sigma_\omega^2)$, $\{\eta_t\} \sim \text{NID}(0, \sigma_\eta^2)$ and $\{\epsilon_t\}, \{\eta_t\}, \{\omega_t\}$ are mutually uncorrelated. The stationary form of (2.3.4) is:

$$\Delta_s y_t = S(L)\eta_t + \Delta_s \gamma_t + \Delta_s \epsilon_t$$

Let a matrix with subscript $[\gamma]$ denote the part of the state-space system matrices (2.2.1), which corresponds to the seasonal component. Then for the dummy seasonality model $\mathbf{Z}_{[\gamma]} = [1, 0, \dots, 0]$ is an $1 \times (s-1)$ vector and,

$$\mathbf{T}_{[\gamma]} = \begin{bmatrix} -1 & -1 & \dots & -1 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & 1 \end{bmatrix}$$

is an $(s-1) \times (s-1)$ matrix.

2.3.2 Trigonometric seasonal model (TS)

In the trigonometric case, the seasonal effect is the combination of $[s/2]$ cycles (Harvey 1989; Priestley 1981) that is $\gamma_t = \sum_{j=1}^{[s/2]} \gamma_{j,t}$ where:

$$\begin{pmatrix} \gamma_{j,t} \\ \gamma_{j,t}^* \end{pmatrix} = \begin{pmatrix} \cos \lambda_j & \sin \lambda_j \\ -\sin \lambda_j & \cos \lambda_j \end{pmatrix} \begin{pmatrix} \gamma_{j,t-1} \\ \gamma_{j,t-1}^* \end{pmatrix} + \begin{pmatrix} \omega_{j,t} \\ \omega_{j,t}^* \end{pmatrix} \quad (2.3.5)$$

where $j = 1, \dots, [s/2]$, $t = 1, \dots, n$ and $\lambda_j = 2\pi j/s$ is the frequency, in radians. The component $\gamma_{j,t}^*$ appear as a matter of construction. The noise terms $\{\omega_{j,t}\} \sim \text{NID}(0, \sigma_j^2)$ and $\{\omega_{j,t}^*\} \sim \text{NID}(0, \sigma_j^2)$ are mutually uncorrelated and $[s/2]$ denotes the integer part of $s/2$

For illustration consider the quarterly case where we have 2 seasonal frequencies $\pi/2$ and π . Then (2.3.5) becomes:

$$\gamma_{1,t} = \gamma_{1,t-1}^* + \omega_{1,t} \quad (2.3.6)$$

$$\gamma_{1,t}^* = -\gamma_{1,t-1} + \omega_{1,t-1}^* \quad (2.3.7)$$

$$\gamma_{2,t} = -\gamma_{2,t-1} + \omega_{2,t} \implies \gamma_{2,t} + \gamma_{2,t-1} = \omega_{2,t} \quad (2.3.8)$$

From the first two equations we get:

$$\gamma_{1,t} = -\gamma_{1,t-2} + \omega_{1,t} + \omega_{1,t-1}^* \implies \gamma_{1,t} + \gamma_{1,t-2} = \omega_{1,t} + \omega_{1,t-1}^* \quad (2.3.9)$$

From (2.3.8) and (2.3.9) we get:

$$S(L)\gamma_{1,t} = \omega_{1,t} + \omega_{1,t-1} + \omega_{1,t-1}^* + \omega_{1,t-2}^* \quad (2.3.10)$$

$$S(L)\gamma_{2,t} = \omega_{2,t} + \omega_{2,t-2} \quad (2.3.11)$$

Thus,

$$S(L)\gamma_t = \sum_{j=1}^2 S(L)\gamma_{j,t} = \omega_{1,t} + \omega_{1,t-1} + \omega_{1,t-1}^* + \omega_{1,t-2}^* + \omega_{2,t} + \omega_{2,t-2} \quad (2.3.12)$$

From (2.3.12) we can see that $\{S(L)\gamma_t\}$ follows an MA(2) process in the quarterly case. Harvey (1989, Ch.2) says that the trigonometric seasonal model is a MA($s-2$) process, without giving a proof. In the next theorem we give a complete proof.

Theorem 2.3.1. $\{S(L)\gamma_t\}$ with γ_t following a trigonometric seasonal model defined in (2.3.16) follows a MA($s-2$) process.

Proof. From (2.3.5) we have:

$$\begin{aligned} \begin{pmatrix} \gamma_{j,t} \\ \gamma_{j,t}^* \end{pmatrix} &= \begin{pmatrix} 1 - (\cos \lambda_j)L & -(\sin \lambda_j)L \\ (\sin \lambda_j)L & 1 - (\cos \lambda_j)L \end{pmatrix}^{-1} \begin{pmatrix} \omega_{j,t} \\ \omega_{j,t}^* \end{pmatrix} \\ \begin{pmatrix} \gamma_{j,t} \\ \gamma_{j,t}^* \end{pmatrix} &= \frac{1}{\delta_j(L)} \begin{pmatrix} 1 - (\cos \lambda_j)L & (\sin \lambda_j)L \\ -(\sin \lambda_j)L & 1 - (\cos \lambda_j)L \end{pmatrix} \begin{pmatrix} \omega_{j,t} \\ \omega_{j,t}^* \end{pmatrix} \\ \implies \delta_j(L)\gamma_{j,t} &= \omega_{j,t} - \omega_{j,t-1} \cos \lambda_j + \omega_{j,t-1}^* \sin \lambda_j \end{aligned}$$

where

$$\delta_j(L) = \begin{cases} 1 - 2 \cos \lambda_j L + L^2 & j = 1, \dots, [s/2] - 1 \\ 1 + L & j = [s/2] \end{cases} \quad (2.3.13)$$

is the trigonometric operator (Harvey 1989, p.21). We put the right hand side of (2.3.13) in MA(1) form by re-writing the error terms and we get:

$$\delta_j(L)\gamma_{j,t} = (1 - \theta_j L)\epsilon_{j,t}, \quad \{\epsilon_{j,t}\} \sim \text{NID}(0, \sigma_j^2) \quad (2.3.14)$$

Using (2.3.14) we can easily see that:

$$S(L)\gamma_t = \sum_{j=1}^{\lfloor s/2 \rfloor} \frac{S(L)}{\delta_j(L)} (1 - \theta_j L)\epsilon_{j,t}$$

For $j = 1, \dots, \lfloor s/2 \rfloor - 1$ the nominator in the right hand side has a lag operator polynomial of order s , while the denominator is of order 2. For $j = \lfloor s/2 \rfloor$ the nominator is of order $s - 1$ and the denominator is of order 1. Thus, the right hand side is a sum of $\lfloor s/2 \rfloor$ independent MA($s-2$) processes and consequently $S(L)\gamma_t$ is itself an MA($s-2$) process. \square

Following the same notation as in §2.3.1, the relevant parts of the system matrices for the trigonometric seasonality model are: $\mathbf{Z}_{[\gamma]} = [1, 1, \dots, 1]$ is an $1 \times (s - 1)$ vector and,

$$\mathbf{T}_{[\gamma]} = \begin{bmatrix} \cos \lambda_1 & \sin \lambda_1 & \dots & 0 \\ -\sin \lambda_1 & \cos \lambda_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & -1 \end{bmatrix}$$

is an $(s - 1) \times (s - 1)$ matrix.

Alternatively we can formulate the seasonal model in terms of $s - 1$ effects associated with the amplitude of deterministic sine and cosine waves defined in the seasonal frequencies $\lambda_j = 2\pi j/s$ for $j = 1, \dots, \lfloor s/2 \rfloor$ (Hannan, Terrell,

and Tuckwell 1970). If these effects are collected in the $(s - 1) \times 1$ vector $\boldsymbol{\tau}_t$, we write (Proietti 2000):

$$\begin{aligned}\gamma_t &= \mathbf{z}'_t \boldsymbol{\tau}_t & (2.3.15) \\ \mathbf{z}'_t &= [\cos \lambda_1 t, \sin \lambda_1 t, \dots, \cos \lambda_{[s/2]} t] \\ \boldsymbol{\tau}_t &= \boldsymbol{\tau}_{t-1} + \boldsymbol{\kappa}_t, \quad \boldsymbol{\kappa}_t \sim \text{NID}(\mathbf{0}_{s-1}, \mathbf{K})\end{aligned}$$

where \mathbf{K} is a diagonal matrix, such that the diagonal elements vary with the frequency. In the common variance case $\mathbf{K} = \sigma_\omega^2 \mathbf{I}_{s-1}$

2.3.3 Harrison and Stevens seasonal model (HS)

An alternative seasonal specification is the Harrison and Stevens (1976) seasonal model. This representation has a time-varying observation equation, in which the seasonal factors are explicitly modelled as a multivariate random walk. For the Harrison-Stevens model, $\mathbf{Z}_{[\gamma]}$ is time-varying and ensures that the correct seasonal factor is related to time t and $\mathbf{T}_{[\gamma]} = \mathbf{I}_s$. The state space model for the seasonal factors is :

$$\begin{aligned}\gamma_t &= \mathbf{x}'_t \boldsymbol{\delta}_t & (2.3.16) \\ \boldsymbol{\delta}_t &= \boldsymbol{\delta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim \text{NID}(\mathbf{0}, \boldsymbol{\Omega})\end{aligned}$$

$$\boldsymbol{\Omega} = \sigma_\omega^2 \left[\mathbf{I}_s - \frac{1}{s} \mathbf{i}_s \mathbf{i}'_s \right] = \sigma_\omega^2 \begin{pmatrix} 1 - \frac{1}{s} & -\frac{1}{s} & \dots & -\frac{1}{s} \\ -\frac{1}{s} & 1 - \frac{1}{s} & \dots & -\frac{1}{s} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{s} & \dots & \dots & 1 - \frac{1}{s} \end{pmatrix} \quad (2.3.17)$$

where $\boldsymbol{\delta}_t$ is an $s \times 1$ vector containing the seasonal effects, $\boldsymbol{x}'_t = [D_{1t}, \dots, D_{st}]$, with $D_{jt} = 1$ in season j and 0 otherwise, and $\mathbf{i}_s = [1, 1, \dots, 1]'$ is an $s \times 1$ vector. From (2.3.16) and (2.3.17) we get:

$$\left. \begin{array}{l} \text{Var}(\mathbf{i}'_s \boldsymbol{\omega}_t) = \mathbf{i}'_s \text{Var}(\boldsymbol{\omega}_t) \mathbf{i}_s = 0 \\ \text{E}(\mathbf{i}'_s \boldsymbol{\omega}_t) = \mathbf{i}'_s \text{E}(\boldsymbol{\omega}_t) = 0 \end{array} \right\} \Rightarrow \mathbf{i}_s \boldsymbol{\omega}_t = 0$$

since $\mathbf{i}'_s \text{Var}(\boldsymbol{\omega}_t) = \mathbf{0}$ and $\text{E}(\boldsymbol{\omega}_t) = \mathbf{0}$ by construction of $\boldsymbol{\omega}_t$. Then from (2.3.16) we have that $\text{E}(\mathbf{i}'_s \boldsymbol{\delta}_t) = \text{E}(\mathbf{i}'_s \boldsymbol{\delta}_{t-1})$, which for $\mathbf{i}_s \boldsymbol{\delta}_0 = \mathbf{0}$ gives $\mathbf{i}_s \boldsymbol{\delta}_t = \mathbf{0}$. Hence seasonal components sum to zero over seasonal periods.

Following Proietti (1998), we get by repeated substitution in (2.3.16):

$$\begin{aligned} \gamma_t &= \mathbf{x}_t \boldsymbol{\delta}_{t-s+1} + \mathbf{x}_t \boldsymbol{\omega}_{t-s+2} + \dots + \mathbf{x}_t \boldsymbol{\omega}_{t-1} + \mathbf{x}_t \boldsymbol{\omega}_t \\ \gamma_{t-1} &= \mathbf{x}_{t-1} \boldsymbol{\delta}_{t-s+1} + \mathbf{x}_{t-1} \boldsymbol{\omega}_{t-s+2} + \dots + \mathbf{x}_{t-1} \boldsymbol{\omega}_{t-1} \\ &\vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots \\ \gamma_{t-s+2} &= \mathbf{x}_{t-s+2} \boldsymbol{\delta}_{t-s+1} + \mathbf{x}_{t-s+2} \boldsymbol{\omega}_{t-s+2} \\ \gamma_{t-s+1} &= \mathbf{x}_{t-s+1} \boldsymbol{\delta}_{t-s+1} \end{aligned}$$

Then:

$$\begin{aligned} S(L)\gamma_t &= \sum_{j=0}^{s-1} \mathbf{x}'_{t-j} \boldsymbol{\delta}_{t-j} \\ &= \sum_{j=0}^{s-1} \sum_{k=j}^{s-2} \mathbf{x}'_{t-j} \boldsymbol{\omega}_{t-j-k} + \left(\sum_{j=0}^{s-1} \mathbf{x}'_{t-j} \right) \boldsymbol{\delta}_{t-s+1} \\ &= \mathbf{x}_t \boldsymbol{\omega}_t + (\mathbf{x}_t + \mathbf{x}_{t-1})' \boldsymbol{\omega}_{t-1} + \dots + (\mathbf{x}_t + \dots + \mathbf{x}_{t-s+2})' \boldsymbol{\omega}_{t-s+2} \end{aligned} \quad (2.3.18)$$

$$\text{since } \mathbf{i}'_s \boldsymbol{\delta}_{t-s+1} = \left(\sum_{j=0}^{s-1} \mathbf{x}'_{t-j} \right) \boldsymbol{\delta}_{t-s+1} = \mathbf{0}.$$

We then prove the following:

Theorem 2.3.2. *{ $S(L)\gamma_t$ } with γ_t following a Harrison and Stevens seasonal model defined in (2.3.16) follows a MA($s-2$) process.*

Proof. From (2.3.18), we see that:

$$\begin{aligned}\text{Var}(S(L)\gamma_t) &= \mathbf{x}'_t \text{Var}(\boldsymbol{\omega}_t) \mathbf{x}_t + \dots + (\mathbf{x}_t + \dots + \mathbf{x}_{t-s+2})' \text{Var}(\boldsymbol{\omega}_t) (\mathbf{x}_t + \dots + \mathbf{x}_{t-s+2}) \\ &= \sum_{k=0}^{s-2} \sum_{j=0}^k \sum_{\ell=0}^k \mathbf{x}'_{t-j} \boldsymbol{\Omega} \mathbf{x}_{t-\ell}\end{aligned}$$

In general the autocovariance function of $S(L)\gamma_t$ is:

$$c(\tau) = \begin{cases} \sum_{k=\tau}^{s-2} \sum_{j=0}^k \sum_{\ell=\tau}^k \mathbf{x}'_{t-j} \boldsymbol{\Omega} \mathbf{x}_{t-\ell} & \text{for } \tau \leq s-2 \\ 0 & \text{for } \tau > s-2 \end{cases}$$

which shows that $S(L)\gamma_t \sim MA(s-2)$. □

2.4 Kalman filter and smoother

This section gives the Kalman filter (Kalman 1960) and smoother equations for the case where the initial state $\boldsymbol{\alpha}_1 \sim N(\mathbf{a}_1, \mathbf{P}_1)$ where \mathbf{a}_1 and \mathbf{P}_1 are known. The Kalman filter and smoother can be modified for the case where some of the elements of \mathbf{a}_1 and \mathbf{P}_1 are unknown. Filtering updates the system each time a new observation comes in. There are several way to derive the Kalman filter, see Anderson and Moore (1979) and references therein; we derive it for the Gaussian case following Durbin and Koopman (2001). The two following simple lemmas from multivariate normal regression theory provide the basis for the treatment of the Kalman filter and smoother in this thesis.

Lemma 2.4.1. Let $(\mathbf{x}, \mathbf{y})' \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where:

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{bmatrix}$$

then the distribution of \mathbf{x} conditional on \mathbf{y} is also multivariate normal with mean:

$$\boldsymbol{\mu}_{x|y} = \boldsymbol{\mu}_x - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)$$

and covariance matrix:

$$\boldsymbol{\Sigma}_{xx|y} = \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}$$

Proof. A proof can be found in Anderson (1984) □

Replacing \mathbf{y} with $\begin{pmatrix} \mathbf{y} \\ \mathbf{z} \end{pmatrix}$ in Lemma 2.4.1 we get the following result.

Lemma 2.4.2. Let \mathbf{x} , \mathbf{y} , and \mathbf{z} be jointly multivariate normal with $\boldsymbol{\mu}_z = \mathbf{0}$ and $\boldsymbol{\Sigma}_{yz} = \mathbf{0}$, then the distribution of \mathbf{x} conditional on \mathbf{y} and \mathbf{z} is also multivariate normal with mean

$$\boldsymbol{\mu}_{x|y,z} = \boldsymbol{\mu}_{x|y} - \boldsymbol{\Sigma}_{xz}\boldsymbol{\Sigma}_{zz}^{-1}\mathbf{z}$$

and covariance matrix

$$\boldsymbol{\Sigma}_{xx|y,z} = \boldsymbol{\Sigma}_{xx|y} - \boldsymbol{\Sigma}_{xz}\boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\Sigma}_{xz}$$

Proof. A proof can be found in Durbin and Koopman (2001) □

2.4.1 Derivation of Kalman filter

The Kalman filter obtains the conditional distribution of $\boldsymbol{\alpha}_{t+1}$ given \mathbf{Y}_t for $t = 1, \dots, n$. Assuming all observations are normally distributed, conditional distributions of any subset are also normal and are completely defined by the first two moments. We use the notation $E(\boldsymbol{\alpha}_t | \mathbf{Y}_{t-1}) = \mathbf{a}_t$ and $\text{Var}(\boldsymbol{\alpha}_t | \mathbf{Y}_{t-1}) = \mathbf{P}_t$. The following derivation is based on Durbin and Koopman (2001):

$$\begin{aligned}
 \mathbf{a}_{t+1} &= E(\boldsymbol{\alpha}_{t+1} | \mathbf{Y}_t) \\
 &= E(\mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{R}_t \boldsymbol{\eta}_t | \mathbf{Y}_t) \\
 &= \mathbf{T}_t E(\boldsymbol{\alpha}_t | \mathbf{Y}_t)
 \end{aligned} \tag{2.4.1}$$

and:

$$\begin{aligned}
 \mathbf{P}_{t+1} &= \text{Var}(\boldsymbol{\alpha}_{t+1} | \mathbf{Y}_t) \\
 &= \text{Var}(\mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{R}_t \boldsymbol{\eta}_t | \mathbf{Y}_t) \\
 &= \mathbf{T}_t \text{Var}(\boldsymbol{\alpha}_t | \mathbf{Y}_t) \mathbf{T}_t' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t'
 \end{aligned} \tag{2.4.2}$$

The one-step ahead prediction error of \mathbf{y}_t given \mathbf{Y}_{t-1} is:

$$\begin{aligned}
 \mathbf{v}_t &= \mathbf{y}_t - E(\mathbf{y}_t | \mathbf{Y}_{t-1}) \\
 &= \mathbf{y}_t - E(\mathbf{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\epsilon}_t | \mathbf{Y}_{t-1}) \\
 &= \mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t
 \end{aligned}$$

When \mathbf{Y}_t is fixed, \mathbf{Y}_{t-1} and \mathbf{y}_t are fixed, so \mathbf{Y}_{t-1} and \mathbf{v}_t are fixed. Consequently:

$$\begin{aligned}
 E(\boldsymbol{\alpha}_t | \mathbf{Y}_t) &= E(\boldsymbol{\alpha}_t | \mathbf{Y}_{t-1}, \mathbf{v}_t) \\
 \text{Var}(\boldsymbol{\alpha}_t | \mathbf{Y}_t) &= \text{Var}(\boldsymbol{\alpha}_t | \mathbf{Y}_{t-1}, \mathbf{v}_t)
 \end{aligned}$$

Using Lemma 2.4.1 and 2.4.2 and the fact that $\mathbf{E}(\boldsymbol{\nu}) = \mathbf{0}$ and $\mathbf{E}(\boldsymbol{\nu}_t | \mathbf{Y}_{t-1}) = \mathbf{E}(\mathbf{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\epsilon}_t - \mathbf{Z}_t \mathbf{a}_t | \mathbf{Y}_{t-1}) = \mathbf{0}$ we have:

$$\begin{aligned}
\mathbf{E}(\boldsymbol{\alpha}_t | \mathbf{Y}_t) &= \mathbf{E}(\boldsymbol{\alpha}_t | \mathbf{Y}_{t-1}, \boldsymbol{\nu}_t) \\
&= \mathbf{E}(\boldsymbol{\alpha}_t | \mathbf{Y}_{t-1}) + \text{Cov}(\boldsymbol{\alpha}_t, \boldsymbol{\nu}_t) \text{Var}(\boldsymbol{\nu}_t)^{-1} \boldsymbol{\nu}_t \\
&= \mathbf{a}_t + \text{Cov}(\boldsymbol{\alpha}_t, \boldsymbol{\nu}_t) \text{Var}(\boldsymbol{\nu}_t)^{-1} \boldsymbol{\nu}_t
\end{aligned} \tag{2.4.3}$$

We have:

$$\begin{aligned}
\text{Cov}(\boldsymbol{\alpha}_t, \boldsymbol{\nu}_t) &= \mathbf{E} \{ \mathbf{E}(\boldsymbol{\alpha}_t \boldsymbol{\nu}_t | \mathbf{Y}_{t-1}) \} \\
&= \mathbf{E} [\mathbf{E} \{ \boldsymbol{\alpha}_t (\mathbf{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\epsilon}_t - \mathbf{Z}_t \mathbf{a}_t)' | \mathbf{Y}_{t-1} \}] \\
&= \mathbf{E} \{ \mathbf{E}(\boldsymbol{\alpha}_t \boldsymbol{\alpha}_t' | \mathbf{Y}_{t-1}) \} \mathbf{Z}_t' - \mathbf{E} \{ \mathbf{E}(\boldsymbol{\alpha}_t | \mathbf{Y}_{t-1}) \} \mathbf{a}_t' \mathbf{Z}_t' \\
&= (\mathbf{P}_t + \mathbf{a}_t \mathbf{a}_t') \mathbf{Z}_t' - \mathbf{a}_t \mathbf{a}_t' \mathbf{Z}_t' \\
&= \mathbf{P}_t \mathbf{Z}_t'
\end{aligned} \tag{2.4.4}$$

since $\mathbf{E}(\boldsymbol{\alpha}_t \boldsymbol{\alpha}_t' | \mathbf{Y}_{t-1}) = \mathbf{P}_t + \mathbf{a}_t \mathbf{a}_t'$ and $\mathbf{E}(\boldsymbol{\alpha}_t \boldsymbol{\epsilon}_t) = \mathbf{0}$. We also have:

$$\begin{aligned}
\mathbf{F}_t = \text{Var}(\boldsymbol{\nu}_t | \mathbf{Y}_{t-1}) &= \text{Var}(\mathbf{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\epsilon}_t - \mathbf{Z}_t \mathbf{a}_t | \mathbf{Y}_{t-1}) \\
&= \mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H}_t
\end{aligned} \tag{2.4.5}$$

Substituting (2.4.3), (2.4.4), and (2.4.5), in (2.4.1) gives:

$$\begin{aligned}
\mathbf{a}_{t+1} &= \mathbf{T}_t \mathbf{a}_t + \mathbf{T}_t \mathbf{P}_t \mathbf{Z}_t' \mathbf{F}_t^{-1} \boldsymbol{\nu}_t \\
&= \mathbf{T}_t \mathbf{a}_t + \mathbf{K}_t \boldsymbol{\nu}_t
\end{aligned} \tag{2.4.6}$$

where:

$$\mathbf{K}_t = \mathbf{T}_t \mathbf{P}_t \mathbf{Z}_t' \mathbf{F}_t^{-1} \tag{2.4.7}$$

We also have:

$$\begin{aligned}
\text{Var}(\boldsymbol{\alpha}_t|\mathbf{Y}_t) &= \text{Var}(\boldsymbol{\alpha}_t|\mathbf{Y}_{t-1}, \mathbf{v}_t) \\
&= \text{Var}(\boldsymbol{\alpha}_t|\mathbf{Y}_{t-1}) - \text{Cov}(\boldsymbol{\alpha}_t, \mathbf{v}_t)\text{Var}(\mathbf{v}_t|\mathbf{Y}_{t-1})\text{Cov}(\boldsymbol{\alpha}_t, \mathbf{v}_t)' \\
&= \mathbf{P}_t - \mathbf{P}_t\mathbf{Z}'_t\mathbf{F}_t^{-1}\mathbf{Z}_t\mathbf{P}'_t
\end{aligned} \tag{2.4.8}$$

Substituting (2.4.8) in (2.4.2) we get:

$$\begin{aligned}
\mathbf{P}_{t+1} &= \mathbf{T}_t\mathbf{P}_t\mathbf{T}'_t - \mathbf{T}_t\mathbf{P}_t\mathbf{Z}'_t\mathbf{K}'_t + \mathbf{R}_t\mathbf{Q}_t\mathbf{R}'_t \\
&= \mathbf{T}_t\mathbf{P}_t\mathbf{L}'_t + \mathbf{R}_t\mathbf{Q}_t\mathbf{R}'_t
\end{aligned} \tag{2.4.9}$$

where:

$$\mathbf{L}_t = \mathbf{T}_t - \mathbf{K}_t\mathbf{Z}_t \tag{2.4.10}$$

Collecting equations (2.4.7)-(2.4.10) we have the Kalman filter equations:

$$\begin{aligned}
\mathbf{v}_t &= \mathbf{y}_t - \mathbf{Z}_t\mathbf{a}_t \\
\mathbf{F}_t &= \mathbf{Z}_t\mathbf{P}_t\mathbf{Z}'_t + \mathbf{H}_t \\
\mathbf{K}_t &= \mathbf{T}_t\mathbf{P}_t\mathbf{Z}'_t\mathbf{F}_t^{-1} \\
\mathbf{L}_t &= \mathbf{T}_t - \mathbf{K}_t\mathbf{Z}_t \\
\mathbf{a}_{t+1} &= \mathbf{T}_t\mathbf{a}_t + \mathbf{K}_t\mathbf{v}_t \\
\mathbf{P}_{t+1} &= \mathbf{T}_t\mathbf{P}_t\mathbf{L}'_t + \mathbf{R}_t\mathbf{Q}_t\mathbf{R}'_t
\end{aligned} \tag{2.4.11}$$

for $t = 1, \dots, n$. \mathbf{v}_t with variance \mathbf{F}_t is the one-step forecast error of \mathbf{y}_t given \mathbf{Y}_{t-1} .

The Kalman filter is said to be in a *steady state* if the recursion for \mathbf{P}_{t+1} is time invariant (Harvey 1989), that is, if:

$$\mathbf{P}_{t+1} = \mathbf{P}_t = \bar{\mathbf{P}}$$

The Kalman filter has a steady state if there exists a time-invariant error covariance matrix $\bar{\mathbf{P}}$ which is the solution to the matrix equation, known also as the *Ricatti equations*:

$$\bar{\mathbf{P}} - \mathbf{T}\bar{\mathbf{P}}\mathbf{T}' + \mathbf{T}\bar{\mathbf{P}}\mathbf{Z}'(\mathbf{Z}\bar{\mathbf{P}}\mathbf{Z}' + \mathbf{H})^{-1}\mathbf{Z}\bar{\mathbf{P}}\mathbf{T}' - \mathbf{R}\mathbf{Q}\mathbf{R} = 0$$

The solution is referred as the *steady state solution* of the Kalman filter. Use of the steady state after convergence leads to considerable computational savings since the inversion of \mathbf{F}_t at each point in time is no longer required.

2.4.2 Fixed interval smoother

The estimation of α_t given all the available observations $\mathbf{y}_1, \dots, \mathbf{y}_n$ is done through smoothing. Using a similar argument as in the filtering, the vector \mathbf{Y}_n is fixed when \mathbf{Y}_{t-1} and $\mathbf{v}_t, \dots, \mathbf{v}_n$ are fixed. The following derivation is based on de Jong (1989) and Durbin and Koopman (2001). Using lemma 2.4.2 and since $\mathbf{v}_t, \dots, \mathbf{v}_n$ are mutually uncorrelated we have:

$$\begin{aligned} \hat{\alpha}_t &= \mathbf{E}(\alpha_t | \mathbf{Y}_n) = \mathbf{E}(\alpha_t | \mathbf{Y}_{t-1}, \mathbf{v}_t, \dots, \mathbf{v}_n) \\ &= \mathbf{a}_t + \sum_{j=t}^n \text{Cov}(\alpha_t, \mathbf{v}_j) \mathbf{F}_j^{-1} \mathbf{v}_j \end{aligned} \quad (2.4.12)$$

Since $E(\mathbf{v}_j | \mathbf{Y}_{j-1}) = \mathbf{0}$ then $\text{Cov}(\boldsymbol{\alpha}_t, \mathbf{v}_j) = E\{E(\boldsymbol{\alpha}_t \mathbf{v}_j' | \mathbf{Y}_{t-1})\}$ and as before $E(\boldsymbol{\alpha}_t \boldsymbol{\alpha}_t') = \mathbf{P}_t + \mathbf{a}_t \mathbf{a}_t'$. We then have:

$$\begin{aligned}
E\{E(\boldsymbol{\alpha}_t \mathbf{v}_t' | \mathbf{Y}_{t-1})\} &= \mathbf{P}_t \mathbf{Z}_t' \\
E\{E(\boldsymbol{\alpha}_t \mathbf{v}_{t+1}' | \mathbf{Y}_{t-1})\} &= E[E\{(\boldsymbol{\alpha}_t (\mathbf{Z}_{t+1} \boldsymbol{\alpha}_{t+1} + \boldsymbol{\epsilon}_{t+1} - \mathbf{Z}_{t+1} \mathbf{a}_{t+1}))' | \mathbf{Y}_{t-1}\}] \\
&= E[E\{(\boldsymbol{\alpha}_t (\mathbf{Z}_{t+1} \mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{Z}_{t+1} \mathbf{R}_t \boldsymbol{\eta}_t + \boldsymbol{\epsilon}_{t+1} \\
&\quad - \mathbf{Z}_{t+1} \mathbf{T}_t \mathbf{a}_t - \mathbf{Z}_{t+1} \mathbf{K}_t \mathbf{v}_t)' | \mathbf{Y}_{t-1}\}] \\
&= \mathbf{P}_t \mathbf{T}_t' \mathbf{Z}_{t+1}' + \mathbf{a}_t \mathbf{a}_t' \mathbf{T}_t' \mathbf{Z}_{t+1}' - \mathbf{a}_t \mathbf{a}_t' \mathbf{T}_t' \mathbf{Z}_{t+1}' - \mathbf{P}_t \mathbf{Z}_t' \mathbf{K}_t' \mathbf{Z}_{t+1}' \\
&= \mathbf{P}_t (\mathbf{T}_t - \mathbf{Z}_t \mathbf{K}_t)' \mathbf{Z}_{t+1}' \\
&= \mathbf{P}_t \mathbf{L}_t' \mathbf{Z}_{t+1}' \tag{2.4.13}
\end{aligned}$$

$$E\{E(\boldsymbol{\alpha}_t \mathbf{v}_{t+2}' | \mathbf{Y}_{t-1})\} = \mathbf{P}_t \mathbf{L}_t' \mathbf{L}_{t+1}' \mathbf{Z}_{t+2}'$$

⋮

$$E\{E(\boldsymbol{\alpha}_t \mathbf{v}_n' | \mathbf{Y}_{t-1})\} = \mathbf{P}_t \mathbf{L}_t' \cdots \mathbf{L}_{n-1}' \mathbf{Z}_n'$$

Substituting (2.4.13) in (2.4.12) we have:

$$\begin{aligned}
\hat{\boldsymbol{\alpha}}_n &= \mathbf{a}_n + \mathbf{P}_n \mathbf{Z}_n' \mathbf{F}_n^{-1} = \mathbf{a}_n + \mathbf{P}_n \mathbf{r}_{n-1} \\
\hat{\boldsymbol{\alpha}}_{n-1} &= \mathbf{a}_{n-1} + \mathbf{P}_{n-1} \mathbf{Z}_{n-1}' \mathbf{F}_{n-1}^{-1} \mathbf{v}_{n-1} + \mathbf{P}_{n-1} \mathbf{L}_{n-1}' \mathbf{Z}_n' \mathbf{F}_n^{-1} \mathbf{v}_n \\
&= \mathbf{a}_{n-1} + \mathbf{P}_{n-1} \mathbf{r}_{n-2} \\
&\quad \vdots \\
\hat{\boldsymbol{\alpha}}_t &= \mathbf{a}_t + \mathbf{P}_t \mathbf{Z}_t' \mathbf{F}_t^{-1} \mathbf{v}_t + \mathbf{P}_t \mathbf{L}_t' \mathbf{Z}_{t+1}' \mathbf{F}_{t+1}^{-1} \mathbf{v}_{t+1} \\
&\quad + \cdots + \mathbf{P}_t \mathbf{L}_t' \cdots \mathbf{L}_{n-1}' \mathbf{Z}_n' \mathbf{F}_n^{-1} \mathbf{v}_n \\
&= \mathbf{a}_t + \mathbf{P}_t \mathbf{r}_{t-1} \tag{2.4.14}
\end{aligned}$$

where:

$$\begin{aligned}
\mathbf{r}_{n-1} &= \mathbf{Z}_n \mathbf{F}_n^{-1} \mathbf{v}_n \\
\mathbf{r}_{n-2} &= \mathbf{Z}'_{n-1} \mathbf{F}_{n-1}^{-1} \mathbf{v}_{n-1} + \mathbf{L}'_{n-1} \mathbf{Z}'_n \mathbf{F}_n^{-1} \mathbf{v}_n \\
&= \mathbf{Z}'_{n-1} \mathbf{F}_{n-1}^{-1} \mathbf{v}_{n-1} + \mathbf{L}'_{n-1} \mathbf{r}_{n-1} \\
&\quad \vdots \\
\mathbf{r}_{t-1} &= \mathbf{Z}'_t \mathbf{F}_t^{-1} \mathbf{v}_t + \mathbf{L}'_t \mathbf{Z}'_{t+1} \mathbf{F}_{t+1}^{-1} \mathbf{v}_{t+1} \\
&\quad + \cdots + \mathbf{L}'_t \cdots \mathbf{L}'_{n-1} \mathbf{Z}'_n \mathbf{F}_n^{-1} \mathbf{v}_n \\
&= \mathbf{Z}'_t \mathbf{F}_t^{-1} \mathbf{v}_t + \mathbf{L}'_t \mathbf{r}_t
\end{aligned} \tag{2.4.15}$$

The variance of the smoothed estimates is (de Jong 1991):

$$\begin{aligned}
\mathbf{V}_t = \text{Var}(\boldsymbol{\alpha}_t | \mathbf{Y}_n) &= \text{Var}(\boldsymbol{\alpha}_t | \mathbf{Y}_{t-1}, \mathbf{v}_t, \dots, \mathbf{v}_n) \\
&= \mathbf{P}_t - \sum_{j=t}^n \text{Cov}(\boldsymbol{\alpha}_t, \mathbf{v}_j) \mathbf{F}_j^{-1} \text{Cov}(\boldsymbol{\alpha}_t, \mathbf{v}_j)'
\end{aligned}$$

Using (2.4.13) we have:

$$\begin{aligned}
\mathbf{V}_n &= \mathbf{P}_n - \mathbf{P}_n \mathbf{Z}'_n \mathbf{F}_n^{-1} \mathbf{Z}_n \mathbf{P}'_n = \mathbf{P}_n - \mathbf{P}_n \mathbf{N}_{n-1} \mathbf{P}'_n \\
\mathbf{V}_{n-1} &= \mathbf{P}_{n-1} - \mathbf{P}_{n-1} \mathbf{Z}'_{n-1} \mathbf{F}_{n-1}^{-1} \mathbf{Z}_{n-1} \mathbf{P}'_{n-1} \\
&\quad - \mathbf{P}_{n-1} \mathbf{L}'_{n-1} \mathbf{Z}'_n \mathbf{F}_n^{-1} \mathbf{Z}_n \mathbf{L}_{n-1} \mathbf{P}'_{n-1} \\
&= \mathbf{P}_{n-1} - \mathbf{P}_{n-1} \mathbf{N}_{n-2} \mathbf{P}'_{n-1} \\
&\quad \vdots \\
\mathbf{V}_t &= \mathbf{P}_t - \mathbf{P}_t \mathbf{Z}'_t \mathbf{F}_t^{-1} \mathbf{Z}_t \mathbf{P}'_t - \mathbf{P}_t \mathbf{L}'_t \mathbf{Z}'_{t+1} \mathbf{F}_{t+1}^{-1} \mathbf{Z}_{t+1} \mathbf{L}_t \mathbf{P}'_t \\
&\quad - \cdots - \mathbf{P}_t \mathbf{L}'_t \cdots \mathbf{L}'_{n-1} \mathbf{Z}'_n \mathbf{F}_n^{-1} \mathbf{Z}_n \mathbf{L}_{n-1} \cdots \mathbf{L}_t \mathbf{P}'_t \\
&= \mathbf{P}_t - \mathbf{P}_t \mathbf{N}_{t-1} \mathbf{P}'_t
\end{aligned} \tag{2.4.16}$$

where:

$$\begin{aligned}
\mathbf{N}_{n-1} &= \mathbf{Z}'_n \mathbf{F}_n^{-1} \mathbf{Z}_n \\
\mathbf{N}_{n-2} &= \mathbf{Z}'_{n-1} \mathbf{F}_{n-1}^{-1} \mathbf{Z}_{n-1} + \mathbf{L}'_{n-1} \mathbf{Z}'_n \mathbf{F}_n^{-1} \mathbf{Z}_n \mathbf{L}_{n-1} \\
&= \mathbf{Z}'_{n-1} \mathbf{F}_{n-1}^{-1} \mathbf{Z}_{n-1} + \mathbf{L}'_{n-1} \mathbf{N}_{n-1} \mathbf{L}_{n-1} \\
&\quad \vdots \\
\mathbf{N}_{t-1} &= \mathbf{Z}'_t \mathbf{F}_t^{-1} \mathbf{Z}_t + \mathbf{L}'_t \mathbf{Z}'_{t+1} \mathbf{F}_{t+1}^{-1} \mathbf{Z}_{t+1} \mathbf{L}_t \\
&\quad + \cdots + \mathbf{L}'_t \cdots \mathbf{L}'_{n-1} \mathbf{Z}'_n \mathbf{F}_n^{-1} \mathbf{Z}_n \mathbf{L}_{n-1} \cdots \mathbf{L}_t \\
&= \mathbf{Z}'_t \mathbf{F}_t^{-1} \mathbf{Z}_t + \mathbf{L}'_t \mathbf{N}_t \mathbf{L}_t
\end{aligned} \tag{2.4.17}$$

Collecting (2.4.14), (2.4.15), (2.4.16), and (2.4.17) we have the smoothing recursions which represent the *fixed interval smoother* as proposed by de Jong (1988), de Jong (1989), and Kohn and Ansley (1989):

$$\begin{aligned}
\hat{\boldsymbol{\alpha}}_t &= \mathbf{a}_t + \mathbf{P}_t \mathbf{r}_{t-1} \\
\mathbf{r}_{t-1} &= \mathbf{Z}'_t \mathbf{F}_t^{-1} \mathbf{v}_t + \mathbf{L}'_t \mathbf{r}_t \\
\mathbf{V}_t &= \mathbf{P}_t - \mathbf{P}_t \mathbf{N}_{t-1} \mathbf{P}_t \\
\mathbf{N}_{t-1} &= \mathbf{Z}'_t \mathbf{F}_t^{-1} \mathbf{Z}_t + \mathbf{L}'_t \mathbf{N}_t \mathbf{L}_t
\end{aligned} \tag{2.4.18}$$

for $t = n, \dots, 1$. \mathbf{r}_{t-1} is a weighted sum of innovations \mathbf{v}_j occurring after time $t - 1$.

The algorithm proposed above provides an alternative to the *classical fixed interval smoother* (Anderson and Moore 1979). This is:

$$\begin{aligned}
\hat{\boldsymbol{\alpha}}_t &= \mathbf{a}_{t|t} + \mathbf{P}_{t|t} \mathbf{T}'_t \mathbf{P}_{t+1}^{-1} (\hat{\boldsymbol{\alpha}}_t - \boldsymbol{\alpha}_t) \\
\mathbf{a}_{t|t} &= \mathbf{E}(\boldsymbol{\alpha}_t | \mathbf{Y}_t) = \mathbf{a}_t + \mathbf{P}_t \mathbf{Z}'_t \mathbf{F}_t^{-1} \mathbf{v}_t \\
\mathbf{P}_{t|t} &= \text{Var}(\boldsymbol{\alpha}_t | \mathbf{Y}_t) = \mathbf{P}_t - \mathbf{P}_t \mathbf{Z}'_t \mathbf{F}_t^{-1} \mathbf{Z}_t \mathbf{P}_t
\end{aligned} \tag{2.4.19}$$

for $t = n, \dots, 1$. As pointed by Koopman (1993), (2.4.19) requires the inversion of \mathbf{P}_t , while (2.4.18) requires only the inversion of \mathbf{F}_t . The advantages are:

1. \mathbf{F}_t has usually smaller dimension than \mathbf{P}_t and;
2. \mathbf{F}_t has already been inverted during the filtering process

2.4.3 Disturbance smoother

We also derive recursions for computing the smoothed estimates $\hat{\boldsymbol{\epsilon}}_t = \mathbf{E}(\boldsymbol{\epsilon}_t | \mathbf{Y}_n)$ and $\hat{\boldsymbol{\eta}}_t = \mathbf{E}(\boldsymbol{\eta}_t | \mathbf{Y}_n)$. Following the same approach as before we have:

$$\hat{\boldsymbol{\epsilon}}_t = \mathbf{E}(\boldsymbol{\epsilon}_t | \mathbf{Y}_{t-1}, \mathbf{v}_t, \dots, \mathbf{v}_n) = \sum_{j=t}^n \mathbf{E}\{\mathbf{E}(\boldsymbol{\epsilon}_t \mathbf{v}'_j | \mathbf{Y}_{t-1})\} \mathbf{F}_j^{-1} \mathbf{v}_j \quad (2.4.20)$$

We have (Koopman 1993):

$$\begin{aligned} \mathbf{E}\{\mathbf{E}(\boldsymbol{\epsilon}_t \mathbf{v}'_t | \mathbf{Y}_{t-1})\} &= \mathbf{E}\{\boldsymbol{\epsilon}_t (\mathbf{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\epsilon}_t - \mathbf{Z}_t \mathbf{a}_t)' | \mathbf{Y}_{t-1}\} = \mathbf{H}_t \\ \mathbf{E}\{\mathbf{E}(\boldsymbol{\epsilon}_t \mathbf{v}'_{t+1} | \mathbf{Y}_{t-1})\} &= \mathbf{E}\left[\mathbf{E}\{\boldsymbol{\epsilon}_t (\mathbf{Z}_{t+1} \boldsymbol{\alpha}_{t+1} + \boldsymbol{\epsilon}_{t+1} - \mathbf{Z}_{t+1} \mathbf{a}_{t+1})' | \mathbf{Y}_{t-1}\}\right] \\ &= \mathbf{E}\left[\mathbf{E}\{\boldsymbol{\epsilon}_t (\mathbf{Z}_{t+1} \mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{Z}_{t+1} \mathbf{R}_t \boldsymbol{\eta}_t + \boldsymbol{\epsilon}_{t+1} \right. \\ &\quad \left. - \mathbf{Z}_{t+1} \mathbf{T}_t \mathbf{a}_t - \mathbf{Z}_{t+1} \mathbf{K}_t \mathbf{v}_t)' | \mathbf{Y}_{t-1}\}\right] \\ &= -\mathbf{H}_t \mathbf{K}'_t \mathbf{Z}'_{t+1} \end{aligned}$$

$$\begin{aligned}
\mathbb{E}\{\mathbb{E}(\boldsymbol{\epsilon}_t \mathbf{v}'_{t+2} | \mathbf{Y}_{t-1})\} &= \mathbb{E}\left[\mathbb{E}\{\boldsymbol{\epsilon}_t(\mathbf{Z}_{t+2}\boldsymbol{\alpha}_{t+2} + \boldsymbol{\epsilon}_{t+2} - \mathbf{Z}_{t+2}\mathbf{a}_{t+2})' | \mathbf{Y}_{t-1}\}\right] \\
&= \mathbb{E}\left[\mathbb{E}\{\boldsymbol{\epsilon}_t(\mathbf{Z}_{t+2}\mathbf{T}_{t+1}\boldsymbol{\alpha}_{t+1} + \mathbf{Z}_{t+2}\mathbf{R}_{t+1}\boldsymbol{\eta}_{t+1} + \boldsymbol{\epsilon}_{t+2} \right. \\
&\quad \left. - \mathbf{Z}_{t+2}\mathbf{T}_{t+1}\mathbf{a}_{t+1} - \mathbf{Z}_{t+2}\mathbf{K}_{t+1}\mathbf{v}_{t+1})' | \mathbf{Y}_{t-1}\}\right] \\
&= \mathbb{E}\left[\mathbb{E}\{\boldsymbol{\epsilon}_t(\mathbf{Z}_{t+2}\mathbf{T}_{t+1}\boldsymbol{\alpha}_{t+1} + \mathbf{Z}_{t+2}\mathbf{R}_{t+1}\boldsymbol{\eta}_{t+1} + \boldsymbol{\epsilon}_{t+2} \right. \\
&\quad \left. - \mathbf{Z}_{t+2}\mathbf{T}_{t+1}\mathbf{T}_t\mathbf{a}_t - \mathbf{Z}_{t+2}\mathbf{T}_{t+1}\mathbf{K}_t\mathbf{v}_t - \right. \\
&\quad \left. \mathbf{Z}_{t+2}\mathbf{K}_{t+1}\mathbf{v}_{t+1})' | \mathbf{Y}_{t-1}\}\right] \\
&= -\mathbf{H}_t\mathbf{K}'_t\mathbf{T}_{t+1}\mathbf{Z}_{t+2} - \mathbf{H}_t\mathbf{K}'_t\mathbf{Z}'_{t+1}\mathbf{K}'_{t+1}\mathbf{Z}'_{t+2} \\
&= -\mathbf{H}_t\mathbf{K}'_t\mathbf{L}'_{t+1}\mathbf{Z}'_{t+2} \\
&\quad \vdots \\
\mathbb{E}\{\mathbb{E}(\boldsymbol{\epsilon}_t \mathbf{v}'_n | \mathbf{Y}_{t-1})\} &= -\mathbf{H}_t\mathbf{K}'_t\mathbf{L}'_{t+1} \cdots \mathbf{L}'_{n-1}\mathbf{Z}'_n \tag{2.4.21}
\end{aligned}$$

As in de Jong (1988) and Kohn and Ansley (1989), substituting (2.4.21) into (2.4.20) we get

$$\begin{aligned}
\hat{\boldsymbol{\epsilon}}_t &= \mathbf{H}_t(\mathbf{F}_t^{-1}\mathbf{v}_t - \mathbf{K}'_t\mathbf{Z}'_{t+1}\mathbf{F}_{t+1}^{-1}\mathbf{v}_{t+1} - \mathbf{K}'_t\mathbf{L}'_{t+1}\mathbf{Z}_{t+2}\mathbf{F}_{t+2}^{-1}\mathbf{v}_{t+2} - \cdots \\
&\quad \mathbf{K}'_t\mathbf{L}_{t+1} \cdots \mathbf{L}'_{n-1}\mathbf{Z}'_n\mathbf{F}_n^{-1}\mathbf{v}_n) \\
&= \mathbf{H}_t(\mathbf{F}_t^{-1}\mathbf{v}_t - \mathbf{K}'_t\mathbf{r}_t) \\
&= \mathbf{H}_t\mathbf{u}_t \tag{2.4.22}
\end{aligned}$$

The smoothed estimate of $\boldsymbol{\eta}_t$ is:

$$\hat{\boldsymbol{\eta}}_t = \mathbb{E}(\boldsymbol{\eta}_t | \mathbf{Y}_{t-1}, \mathbf{v}_t, \dots, \mathbf{v}_n) = \sum_{j=t}^n \mathbb{E}\{\mathbb{E}(\boldsymbol{\eta}_t \mathbf{v}'_j | \mathbf{Y}_{t-1})\} \mathbf{F}_j^{-1} \mathbf{v}_j \tag{2.4.23}$$

We have:

$$\begin{aligned}
\mathbb{E}\{\mathbb{E}(\boldsymbol{\eta}_t \mathbf{v}'_t | \mathbf{Y}_{t-1})\} &= \mathbb{E}\{\boldsymbol{\eta}_t (\mathbf{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\epsilon}_t - \mathbf{Z}_t \mathbf{a}_t)' | \mathbf{Y}_{t-1}\} \\
&= \mathbb{E}[\mathbb{E}\{\boldsymbol{\eta}_t (\mathbf{Z}_t \mathbf{T}_{t-1} \boldsymbol{\alpha}_{t-1} + \mathbf{Z}_t \mathbf{R}_{t-1} \boldsymbol{\eta}_{t-1} + \boldsymbol{\epsilon}_t - \mathbf{Z}_t \mathbf{a}_t)' | \mathbf{Y}_{t-1}\}] \\
&= \mathbf{0} \\
\mathbb{E}\{\mathbb{E}(\boldsymbol{\eta}_t \mathbf{v}'_{t+1} | \mathbf{Y}_{t-1})\} &= \mathbb{E}[\mathbb{E}\{\boldsymbol{\eta}_t (\mathbf{Z}_{t+1} \boldsymbol{\alpha}_{t+1} + \boldsymbol{\epsilon}_{t+1} - \mathbf{Z}_{t+1} \mathbf{a}_{t+1})' | \mathbf{Y}_{t-1}\}] \\
&= \mathbb{E}[\mathbb{E}\{\boldsymbol{\eta}_t (\mathbf{Z}_{t+1} \mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{Z}_{t+1} \mathbf{R}_t \boldsymbol{\eta}_t + \boldsymbol{\epsilon}_{t+1} - \mathbf{Z}_{t+1} \mathbf{a}_{t+1})' | \mathbf{Y}_{t-1}\}] \\
&= \mathbf{Q}_t \mathbf{R}'_t \mathbf{Z}'_{t+1} \\
\mathbb{E}\{\mathbb{E}(\boldsymbol{\eta}_t \mathbf{v}'_{t+2} | \mathbf{Y}_{t-1})\} &= \mathbb{E}[\mathbb{E}\{\boldsymbol{\eta}_t (\mathbf{Z}_{t+2} \boldsymbol{\alpha}_{t+2} + \boldsymbol{\epsilon}_{t+2} - \mathbf{Z}_{t+2} \mathbf{a}_{t+2})' | \mathbf{Y}_{t-1}\}] \\
&= \mathbb{E}[\mathbb{E}\{\boldsymbol{\eta}_t (\mathbf{Z}_{t+2} \mathbf{T}_{t+1} \mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{Z}_{t+2} \mathbf{T}_{t+1} \mathbf{R}_t \boldsymbol{\eta}_t \\
&\quad + \mathbf{Z}_{t+2} \mathbf{R}_{t+1} \boldsymbol{\eta}_{t+1} + \boldsymbol{\epsilon}_{t+2} - \mathbf{Z}_{t+2} \mathbf{T}_t \mathbf{a}_{t+1})' | \mathbf{Y}_{t-1}\}] \\
&= \mathbf{Z}_{t+2} \mathbf{K}_{t+1} \mathbf{v}_{t+1})' \\
&= \mathbf{Q}_t \mathbf{R}'_t \mathbf{T}'_{t+1} \mathbf{Z}'_{t+2} - \mathbf{Q}_t \mathbf{R}'_t \mathbf{Z}'_{t+1} \mathbf{K}_{t+1} \mathbf{Z}'_{t+2} \\
&= \mathbf{Q}_t \mathbf{R}'_t \mathbf{L}'_{t+1} \mathbf{Z}'_{t+2} \\
&\quad \vdots
\end{aligned} \tag{2.4.24}$$

$$\mathbb{E}\{\mathbb{E}(\boldsymbol{\eta}_t \mathbf{v}'_n | \mathbf{Y}_{t-1})\} = \mathbf{Q}_t \mathbf{R}'_t \mathbf{L}'_{t+1} \cdots \mathbf{L}'_{n-1} \mathbf{Z}'_n \tag{2.4.25}$$

As in Koopman (1993), substituting (2.4.24) into (2.4.23) we get

$$\begin{aligned}
\hat{\boldsymbol{\eta}}_t &= \mathbf{Q}_t \mathbf{R}'_t (\mathbf{Z}'_{t+1} \mathbf{F}_{t+1}^{-1} \mathbf{v}_{t+1} + \mathbf{L}'_{t+1} \mathbf{Z}'_{t+2} \mathbf{F}_{t+2}^{-1} \mathbf{v}_{t+2} - \cdots \\
&\quad \mathbf{L}'_{t+1} \cdots \mathbf{L}'_{n-1} \mathbf{Z}'_n \mathbf{F}_n^{-1} \mathbf{v}_n) \\
&= \mathbf{Q}_t \mathbf{R}'_t \mathbf{r}_t \tag{2.4.26}
\end{aligned}$$

We also derive the variances of the smoothed disturbances. Using lemmas (2.4.1) and (2.4.2) and equations (2.4.21) we get (de Jong 1988; Kohn and

Ansley 1989)

$$\begin{aligned}
\text{Var}(\boldsymbol{\epsilon}_t | \mathbf{Y}_n) &= \text{Var}(\boldsymbol{\epsilon}_t | \mathbf{Y}_{t-1}, \mathbf{v}_t, \dots, \mathbf{v}_n) \\
&= \text{Var}(\boldsymbol{\epsilon}_t | \mathbf{Y}_{t-1} - \sum_{j=t}^n \text{Cov}(\boldsymbol{\epsilon}_t, \mathbf{v}_j) \text{Var}(\mathbf{v}_j | \mathbf{Y}_{j-1})^{-1} \text{Cov}(\boldsymbol{\epsilon}_t, \mathbf{v}_j)') \\
&= \mathbf{H}_t - \sum_{j=t}^n \mathbf{E}\{\mathbf{E}(\boldsymbol{\epsilon}_t \mathbf{v}_j' | \mathbf{Y}_{t-1})\} \mathbf{F}_j^{-1} \mathbf{E}\{\mathbf{E}(\boldsymbol{\epsilon}_t \mathbf{v}_j | \mathbf{Y}_{t-1})\}' \\
&= \mathbf{H}_t - \mathbf{H}_t (\mathbf{F}_t^{-1} + \mathbf{K}_t' \mathbf{Z}_{t+1} \mathbf{F}_{t+1}^{-1} \mathbf{Z}_{t+1} \mathbf{K}_t \\
&\quad \mathbf{K}_t' \mathbf{L}'_{t+1} \mathbf{Z}'_{t+2} \mathbf{F}_{t+2}^{-1} \mathbf{Z}_{t+2} \mathbf{L}_{t+1} \mathbf{K}_t - \dots \\
&\quad - \mathbf{K}_t' \mathbf{L}'_{t+1} \dots \mathbf{L}'_{n-1} \mathbf{Z}'_n \mathbf{F}_n^{-1} \mathbf{Z}_n \mathbf{L}_{n-1} \dots \mathbf{L}_{t+1} \mathbf{K}_t) \mathbf{H}_t' \\
&= \mathbf{H}_t - \mathbf{H}_t (\mathbf{F}_t^{-1} + \mathbf{K}_t' \mathbf{N}_t \mathbf{K}_t) \mathbf{H}_t' \\
&= \mathbf{H}_t - \mathbf{H}_t \mathbf{D}_t \mathbf{H}_t' \tag{2.4.27}
\end{aligned}$$

In a similar way, using equations (2.4.24) we get: (Koopman 1993)

$$\begin{aligned}
\text{Var}(\boldsymbol{\eta}_t | \mathbf{Y}_n) &= \text{Var}(\boldsymbol{\eta}_t | \mathbf{Y}_{t-1}, \mathbf{v}_t, \dots, \mathbf{v}_n) \\
&= \text{Var}(\boldsymbol{\eta}_t | \mathbf{Y}_{t-1} - \sum_{j=t}^n \text{Cov}(\boldsymbol{\eta}_t, \mathbf{v}_j) \text{Var}(\mathbf{v}_j | \mathbf{Y}_{j-1})^{-1} \text{Cov}(\boldsymbol{\eta}_t, \mathbf{v}_j)') \\
&= \mathbf{H}_t - \sum_{j=t}^n \mathbf{E}\{\mathbf{E}(\boldsymbol{\eta}_t \mathbf{v}_j' | \mathbf{Y}_{t-1})\} \mathbf{F}_j^{-1} \mathbf{E}\{\mathbf{E}(\boldsymbol{\eta}_t \mathbf{v}_j | \mathbf{Y}_{t-1})\}' \\
&= \mathbf{Q}_t - \mathbf{Q}_t \mathbf{R}_t' (\mathbf{Z}_{t+1} \mathbf{F}_{t+1}^{-1} \mathbf{Z}_{t+1} + \mathbf{L}'_{t+1} \mathbf{Z}'_{t+2} \mathbf{F}_{t+2}^{-1} \mathbf{Z}_{t+2} \mathbf{L}_{t+1} + \dots \\
&\quad + \mathbf{L}'_{t+1} \dots \mathbf{L}'_{n-1} \mathbf{Z}'_n \mathbf{F}_n^{-1} \mathbf{Z}_n \mathbf{L}_{n-1} \dots \mathbf{L}_{t+1}) \mathbf{R}_t \mathbf{Q}_t' \\
&= \mathbf{Q}_t - \mathbf{Q}_t \mathbf{R}_t' \mathbf{N}_t \mathbf{R}_t \mathbf{Q}_t' \tag{2.4.28}
\end{aligned}$$

Collecting (2.4.22), (2.4.26), (2.4.27), and (2.4.28) we have the *disturbance smoothing recursions* (de Jong 1988; Koopman 1993):

$$\begin{aligned}
\hat{\boldsymbol{\epsilon}}_t &= \mathbf{H}_t \mathbf{u}_t \\
\mathbf{u}_t &= \mathbf{F}_t^{-1} \mathbf{v}_t - \mathbf{K}'_t \mathbf{r}_t \\
\text{Var}(\boldsymbol{\epsilon}_t | \mathbf{Y}_n) &= \mathbf{H}_t \mathbf{H}_t \mathbf{D}_t \mathbf{H}_t' \\
\mathbf{D}_t &= \mathbf{F}_t^{-1} + \mathbf{K}'_t \mathbf{N}_t \mathbf{K}_t \\
\hat{\boldsymbol{\eta}}_t &= \mathbf{Q}_t \mathbf{R}'_t \mathbf{r}_t \\
\text{Var}(\boldsymbol{\eta}_t | \mathbf{Y}_n) &= \mathbf{Q}_t - \mathbf{Q}_t \mathbf{R}'_t \mathbf{N}_t \mathbf{R}_t \mathbf{Q}_t' \tag{2.4.29}
\end{aligned}$$

where \mathbf{r}_t and \mathbf{N}_t are derived from recursions (2.4.18)

2.5 Square root filter

In this section, we present the equations for the square root filter (Morf and Kailath 1975). Seasonal models are particularly susceptible to round-off errors that may result in a negative definite value for the conditional state covariance matrix \mathbf{P}_t . The problem is avoided by using the square root filter. This filter is based on orthogonal lower triangular transformations for which we use Givens rotation techniques (Golub and van Loan 1996).

2.5.1 Givens rotations

Let \mathbf{U} be a $m \times n$ matrix with $m \geq n$. We would like to transform \mathbf{U} to an upper triangular matrix \mathbf{U}^* using orthogonal matrix \mathbf{G} , such that $\mathbf{G}\mathbf{G}' = \mathbf{I}_m$. We define the Givens matrix $\mathbf{G}(i-1, i, \theta)$ as the identity matrix \mathbf{I}_m with four elements replaced by:

$$\begin{aligned} G_{i,i} = G_{i-1,i-1} &= c \\ G_{i-1,i} &= s \\ G_{i,i-1} &= -s \end{aligned}$$

where $c = \cos(\theta)$ and $s = \sin(\theta)$ for some θ . Premultiplication of \mathbf{U} by $\mathbf{G}(i-1, i, \theta)$ is the same as a counterclockwise rotation of θ radians in the $(i-1, i)$ plane. The element of \mathbf{U}^* in the k^{th} row and l^{th} column is then:

$$U_{k,l}^* = \begin{cases} cU_{i-1,k} - sU_{i,k} & k = i-1 \\ cU_{i-1,k} + sU_{i,k} & k = i \\ U_{j,k} & k \neq i-1, i \end{cases}$$

It is clear we can force $U_{i,l}^*$ to be zero by setting:

$$\begin{aligned} c &= \frac{x_1}{\sqrt{x_1^2 + x_2^2}} \\ s &= -\frac{x_2}{\sqrt{x_1^2 + x_2^2}} \end{aligned}$$

where:

$$\begin{aligned} x_1 &= U_{i-1,l} \\ x_2 &= U_{i,l} \end{aligned}$$

for which $c^2 + s^2 = 1$ and $cx_1 + sx_2 = 0$. Since $\mathbf{G}\mathbf{G}' = \mathbf{I}_m$, Givens rotations can be applied repeatedly to create zero blocks in a matrix with the overall

transformation being orthogonal. It follows that, if $m \leq n$, \mathbf{U} can be transformed to a lower triangular matrix by applying the previous transformations to the transpose \mathbf{U}' . In §A.1 we provide a code written in Ox (Doornik 1998) that applies repeated Givens rotation to create a lower triangular matrix

2.5.2 Square root form

Following the notation of (Durbin and Koopman 2001), we define the partitioned matrix \mathbf{U}_t by:

$$\mathbf{U}_t = \begin{pmatrix} \mathbf{Z}_t \tilde{\mathbf{P}}_t & \tilde{\mathbf{H}}_t & 0 \\ \mathbf{T}_t \tilde{\mathbf{P}}_t & 0 & \mathbf{R}_t \tilde{\mathbf{Q}}_t \end{pmatrix}$$

where $\tilde{\mathbf{P}}_t$, $\tilde{\mathbf{H}}_t$ and $\tilde{\mathbf{Q}}_t$ are lower triangular matrices so that :

$$\begin{aligned} \mathbf{P}_t &= \tilde{\mathbf{P}}_t \tilde{\mathbf{P}}_t' \\ \mathbf{H}_t &= \tilde{\mathbf{H}}_t \tilde{\mathbf{H}}_t' \\ \mathbf{Q}_t &= \tilde{\mathbf{Q}}_t \tilde{\mathbf{Q}}_t' \end{aligned}$$

In Harrison and Stevens seasonal model 2.3.16 \mathbf{Q}_t is not of full rank and the Choleski decomposition cannot be used. Since \mathbf{Q}_t is square, it can be decomposed as:

$$\mathbf{Q}_t = \mathbf{C}_t \mathbf{\Lambda}_t \mathbf{C}_t'$$

where \mathbf{C}_t is a matrix of eigenvectors and $\mathbf{\Lambda}_t$ is a diagonal matrix with the eigenvalues in the diagonal. Then we apply the Givens rotations to the matrix $\mathbf{Q}^{1/2} = \mathbf{C} \mathbf{\Lambda}^{1/2}$ to get $\tilde{\mathbf{Q}}_t$. It follows that:

$$\mathbf{U}_t \mathbf{U}_t' = \begin{pmatrix} \mathbf{F}_t & \mathbf{Z}_t \mathbf{P}_t \mathbf{T}_t \\ \mathbf{T}_t \mathbf{P}_t \mathbf{Z}_t' & \mathbf{T}_t \mathbf{P}_t \mathbf{T}_t' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t' \end{pmatrix}$$

Applying Givens rotations to \mathbf{U}_t to get a lower triangular matrix \mathbf{U}_t^* :

$$\mathbf{U}_t^* = \begin{pmatrix} \mathbf{U}_{1,t}^* & 0 & 0 \\ \mathbf{U}_{2,t}^* & \mathbf{U}_{3,t}^* & 0 \end{pmatrix}$$

It follows that (Morf and Kailath 1975):

$$\begin{aligned} \mathbf{U}_t^* \mathbf{U}_t^{*'} &= \begin{pmatrix} \mathbf{U}_{1,t}^{*'} \mathbf{U}_{1,t}^* & \mathbf{U}_{1,t}^* \mathbf{U}_{2,t}^{*'} \\ \mathbf{U}_{2,t}^* \mathbf{U}_{1,t}^{*'} & \mathbf{U}_{2,t}^* \mathbf{U}_{2,t}^{*'} + \mathbf{U}_{3,t}^* \mathbf{U}_{3,t}^{*'} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{F}_t & \mathbf{Z}_t \mathbf{P}_t \mathbf{T}_t \\ \mathbf{T}_t \mathbf{P}_t \mathbf{Z}_t' & \mathbf{T}_t \mathbf{P}_t \mathbf{T}_t' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t' \end{pmatrix} \end{aligned}$$

so that:

$$\begin{aligned} \mathbf{U}_{1,t}^* &= \tilde{\mathbf{F}}_t \\ \mathbf{U}_{2,t}^* &= \mathbf{T}_t \mathbf{P}_t \mathbf{Z}_t' \tilde{\mathbf{F}}_t^{-1} = \mathbf{K}_t \tilde{\mathbf{F}}_t \\ \mathbf{U}_{3,t}^* &= \tilde{\mathbf{P}}_{t+1} \end{aligned}$$

$\tilde{\mathbf{P}}_{t+1}$ is then used to give an update for \mathbf{U}_t . The update for the state vector \mathbf{a}_t is (Durbin and Koopman 2001):

$$\begin{aligned} \mathbf{a}_{t+1} &= \mathbf{T}_t \mathbf{a}_t + \mathbf{K}_t \mathbf{v}_t \\ &= \mathbf{T}_t \mathbf{a}_t + \mathbf{T}_t \mathbf{P}_t \mathbf{Z}_t' \tilde{\mathbf{F}}_t^{-1} \mathbf{v}_t \\ &= \mathbf{T}_t \mathbf{a}_t \mathbf{U}_{2,t}^* \mathbf{U}_{1,t}^{*-1} \mathbf{v}_t \end{aligned}$$

where $\mathbf{v}_t = \mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t$.

2.6 Initialisation

In order to start the filtering and smoothing operations, we need to make certain assumption for the distribution of $\boldsymbol{\alpha}_1$. The variance matrix \mathbf{P}_1 of

the initial vector \mathbf{a}_1 contains diffuse elements when some components of the state are non-stationary, see Ansley and Kohn (1985), de Jong (1991), and Koopman (1997):

$$\mathbf{P}_1 = \mathbf{P}_{*,1} + \kappa \mathbf{P}_{\infty,1}$$

This formulation implies that some Kalman filter quantities are also diffuse. The exact initial Kalman filter is then (Koopman 1997):

$$\begin{aligned}
\mathbf{a}_{t+1} &= \mathbf{T}_t \mathbf{a}_t + \mathbf{K}_{*,t} \mathbf{v}_t \\
\mathbf{P}_{*,t+1} &= \mathbf{T}_t \mathbf{P}_{*,t} \mathbf{T}_t' - \mathbf{C}_{*,t} + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t' \\
\mathbf{P}_{\infty,t+1} &= \mathbf{T}_t \mathbf{P}_{\infty,t} \mathbf{T}_t' - \mathbf{C}_{\infty,t} \\
\mathbf{K}_{*,t} &= \mathbf{M}_{*,t} \mathbf{F}_{*,t}^- + \mathbf{M}_{\infty,t} \mathbf{F}_{\infty,t}^- \\
\mathbf{F}_{*,t} &= \mathbf{Z}_t \mathbf{P}_{*,t} \mathbf{Z}_t' + \mathbf{H}_t \\
\mathbf{F}_{\infty,t} &= \mathbf{Z}_t \mathbf{P}_{\infty,t} \mathbf{Z}_t' \\
\mathbf{M}_{*,t} &= \mathbf{T}_t \mathbf{P}_{*,t} \mathbf{Z}_t' \\
\mathbf{M}_{\infty,t} &= \mathbf{T}_t \mathbf{P}_{\infty,t} \mathbf{Z}_t' \\
\mathbf{C}_{*,t} &= \mathbf{M}_{*,t} \mathbf{K}_{*,t} + \mathbf{M}_{\infty,t} \mathbf{F}_{\infty,t}^- (\mathbf{M}_{*,t} - \mathbf{M}_{\infty,t} \mathbf{F}_{\infty,t}^- \mathbf{F}_{*,t}^-)' \\
\mathbf{C}_{\infty,t} &= \mathbf{M}_{\infty,t} \mathbf{F}_{\infty,t}^- \mathbf{M}_{\infty,t}'
\end{aligned} \tag{2.6.1}$$

$\mathbf{F}_{*,t}^-$ and $\mathbf{F}_{\infty,t}^-$ are calculated by the diagonalising $\mathbf{F}_{*,t}$ and $\mathbf{F}_{\infty,t}$ in the following way:

$$(\mathbf{J}_{1,t}, \mathbf{J}_{2,t})' \mathbf{F}_{\infty,t} (\mathbf{J}_{1,t}, \mathbf{J}_{2,t}) = \begin{bmatrix} \mathbf{I}_r & 0 \\ 0 & 0 \end{bmatrix} (\mathbf{J}_{1,t}, \mathbf{J}_{2,t})' \mathbf{F}_{*,t} (\mathbf{J}_{1,t}, \mathbf{J}_{2,t}) = \begin{bmatrix} \mathbf{V}_{*,t} & 0 \\ 0 & \mathbf{I}_{N-r} \end{bmatrix}$$

where $\mathbf{J}_t = (\mathbf{J}_{1,t}, \mathbf{J}_{2,t})$ is a nonsingular matrix, $r = \text{rank}(\mathbf{F}_{\infty,t})$ and $\text{rank}(\mathbf{V}_{*,t}) \leq r$. Then $\mathbf{F}_{\infty,t}^- = \mathbf{J}_{1,t} \mathbf{J}_{1,t}'$ and $\mathbf{F}_{*,t}^- = \mathbf{J}_{2,t} \mathbf{J}_{2,t}'$. The exact initial Kalman filter starts off with $\mathbf{a}_1 = \mathbf{0}$, $\mathbf{P}_{*,1} = \mathbf{0}$ and $\mathbf{P}_{\infty,1} = \mathbf{I}_m$

2.7 Estimation

We use Gaussian maximum likelihood for estimation and inference. When the initial conditions are known, so that $\boldsymbol{\alpha}_1$ has density $N(\boldsymbol{\alpha}_1, \mathbf{P}_1)$ where $\boldsymbol{\alpha}_1$ and \mathbf{P}_1 are known then the log-likelihood is (Harvey 1993):

$$\log L(\boldsymbol{\psi}; \mathbf{Y}_n) = \sum_{t=1}^n \log f(\mathbf{y}_t | \mathbf{Y}_{t-1}) \quad (2.7.1)$$

where $\boldsymbol{\psi}$ is the vector of parameters, $f(\mathbf{y}_t | \mathbf{Y}_{t-1})$ is the conditional density function of \mathbf{y}_t and $f(\mathbf{y}_1 | \mathbf{Y}_0) = f(\mathbf{y}_1)$. For the structural models, $\boldsymbol{\psi}$ is usually a vector of the variance parameters that need to be estimated. Since $E(\mathbf{y}_t | \mathbf{Y}_{t-1}) = \mathbf{Z}_t \mathbf{a}_t$, $\mathbf{v}_t = \mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t$ and $\mathbf{F}_t = \text{Var}(\mathbf{y}_t | \mathbf{Y}_{t-1})$, equation (2.7.1) becomes:

$$\log L(\boldsymbol{\psi}; \mathbf{Y}_n) = -\frac{np}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^n (\log |\mathbf{F}_t| + \mathbf{v}_t' \mathbf{F}_t^{-1} \mathbf{v}_t) \quad (2.7.2)$$

Given the parameter values, the likelihood can be evaluated using a single run of the Kalman filter. The representation in (2.7.2) is often referred to as the *prediction error decomposition* of the likelihood function (Harvey 1989).

2.7.1 Numerical optimisation algorithms

As we saw in, the likelihood function (2.7.2) is a function of a vector $\boldsymbol{\psi}$ of unknown parameters. We estimate $\boldsymbol{\psi}$ by maximising the likelihood function by iterative numerical procedures. The most widely used numerical procedure for optimising a function is Newton's method. The basis for Newton's

method is a linear Taylor series approximation. It solves the following equation (Durbin and Koopman 2001):

$$\partial_1(\boldsymbol{\psi}) = \frac{\partial \log L(\boldsymbol{\psi}; \mathbf{Y}_n)}{\partial \boldsymbol{\psi}} = 0 \quad (2.7.3)$$

Using the first-order Taylor series around an arbitrary point $\tilde{\boldsymbol{\psi}}$ yields:

$$\partial_1(\boldsymbol{\psi}) = \partial_1(\boldsymbol{\psi})|_{\boldsymbol{\psi}=\tilde{\boldsymbol{\psi}}} + \partial_2(\boldsymbol{\psi})|_{\boldsymbol{\psi}=\tilde{\boldsymbol{\psi}}}(\boldsymbol{\psi} - \tilde{\boldsymbol{\psi}}) \quad (2.7.4)$$

where:

$$\partial_2(\boldsymbol{\psi}) = \frac{\partial^2 \log L(\boldsymbol{\psi}; \mathbf{Y}_n)}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'}$$

Using (2.7.3) and (2.7.4) and then equating $\boldsymbol{\psi}$ to $\boldsymbol{\psi}_{i+1}$ and $\tilde{\boldsymbol{\psi}}$ to $\boldsymbol{\psi}_i$ we obtain the following iteration:

$$\boldsymbol{\psi}_{i+1} = \boldsymbol{\psi}_i - \partial_2(\boldsymbol{\psi}_i)^{-1} \partial_1(\boldsymbol{\psi}_i) \quad (2.7.5)$$

The iteration is repeated until it converges or until a switch is made to another optimisation method. Newton's method will converge very rapidly in many situations. If the Hessian matrix $\partial_2(\boldsymbol{\psi})$ is negative definite for all $\boldsymbol{\psi}$ then a unique maximum exists for the likelihood function. The first derivative $\partial_1(\boldsymbol{\psi})$ gives the direction of the step taken to the optimum and $\partial_2(\boldsymbol{\psi})$ modifies the size of the step. We can modify (2.7.5):

$$\boldsymbol{\psi}_{i+1} = \boldsymbol{\psi}_i + s \partial_2(\boldsymbol{\psi}_i)^{-1} \partial_1(\boldsymbol{\psi}_i) \quad (2.7.6)$$

by including a line search within the optimisation process for s . The optimal value for s is usually found to be between 0 and 1.

In practice it is often difficult to compute $\partial_1(\boldsymbol{\psi})$ and $\partial_2(\boldsymbol{\psi})$ analytically. The programming language Ox (Doornik 1998) implements the quasi-Newton

method developed by Broyden, Fletcher, Golfarb, and Shanno (BFGS) (Fletcher 1987). This method uses supplied analytical or numerical first derivatives. If analytical derivatives are not provided, then at each iteration for $\boldsymbol{\psi}$, a value for $\partial_2(\boldsymbol{\psi})^{-1}$ is obtained by the following recursion:

$$\partial_2(\boldsymbol{\psi}_{i+1})^{-1} = \partial_2(\boldsymbol{\psi}_i)^{-1} + \left(s + \frac{g_i' g_i^*}{\pi \boldsymbol{\psi}_i' g_i} \right) \frac{\pi(\boldsymbol{\psi}_i) \pi(\boldsymbol{\psi}_i)'}{\boldsymbol{\psi}(\boldsymbol{\psi}_i)' g_i} - \frac{\pi(\boldsymbol{\psi}_i) g_i^{*'} + g_i^* \pi(\boldsymbol{\psi}_i)'}{\pi(\boldsymbol{\psi}_i)' g_i}$$

where:

$$\begin{aligned} g_i &= \partial_1(\boldsymbol{\psi}_i) - \partial_1(\boldsymbol{\psi}_{i-1}) \\ g_i^* &= \partial_2(\boldsymbol{\psi}_i)^{-1} g_i \\ \pi(\boldsymbol{\psi}_i) &= -\partial_2(\boldsymbol{\psi}_i)^{-1} \partial_1(\boldsymbol{\psi}_i) \end{aligned}$$

For details and derivations of the Newton's method and the BFGS methods see Fletcher (1987).

2.7.2 The score vector

The procedures described in §2.7.1 usually update the parameters at each run by using the gradient or score vector:

$$\partial_1(\boldsymbol{\psi}) = \frac{\partial \log L(\boldsymbol{\psi}; \mathbf{Y}_n)}{\partial \boldsymbol{\psi}}$$

Following Koopman and Shephard (1992) we derive a general expression for the score vectors of structural time series models. We assume that the system matrices \mathbf{Z}_t and \mathbf{T}_t have no unknown parameters and only \mathbf{H}_t and \mathbf{Q}_t have unknown parameters. This is the case for all the applications we use in this thesis. Let $f(\boldsymbol{\alpha}, \mathbf{Y}_n; \boldsymbol{\psi})$ be the joint density of $\boldsymbol{\alpha}$ and \mathbf{Y}_n where $\boldsymbol{\alpha} =$

$(\alpha_1, \dots, \alpha_n)$, and ψ is the vector of unknown parameters. Let $f(\alpha|\mathbf{Y}_n; \psi)$ be the conditional density of α given \mathbf{Y}_n , $f(\mathbf{Y}_n; \psi)$ be the marginal density of \mathbf{Y}_n , and $f(\alpha; \psi)$ be the marginal density of α . Using known results for conditional densities, we have:

$$\begin{aligned}
\log f(\mathbf{Y}_n; \psi) &= \log f(\mathbf{Y}_n, \alpha; \psi) - \log f(\alpha|\mathbf{Y}_n; \psi) \\
&= \log f(\mathbf{Y}_n|\alpha; \psi) + \log f(\alpha; \psi) - \log f(\alpha|\mathbf{Y}_n; \psi) \\
&= \sum_{t=1}^n [\log f(y_t|\alpha_{t-1}; \psi) + \log f(\alpha_t|\alpha_{t-1}; \psi)] \\
&\quad + \log f(\alpha_0; \psi) - \log f(\alpha|\mathbf{Y}_n; \psi) \\
&= -\frac{1}{2} \sum_{t=1}^n (\log |\mathbf{H}(\psi)| + \text{tr}[\mathbf{H}_t(\psi)^{-1}(y_t - \mathbf{Z}_t\alpha_t)(y_t - \mathbf{Z}_t\alpha_t)']) \\
&\quad -\frac{1}{2} \sum_{t=1}^n (\log |\mathbf{Q}(\psi)| + \text{tr}[\mathbf{Q}_t(\psi)^{-1}(\alpha_t - \mathbf{T}_t\alpha_{t-1})(\alpha_t - \mathbf{T}_t\alpha_{t-1})']) \\
&\quad -\frac{1}{2} \log |\mathbf{P}_0| - \frac{1}{2}(\alpha_0 - \mathbf{a}_0)'\mathbf{P}_0^{-1}(\alpha_0 - \mathbf{a}_0) - \log f(\alpha|\mathbf{Y}_n; \psi)
\end{aligned}$$

In order to derive the score at a point ψ^* , we first integrate both sides with respect to $f(\alpha|\mathbf{Y}_n; \psi^*)$ and then differentiate with respect to ψ :

$$\begin{aligned}
\left. \frac{\partial \log f(\mathbf{Y}_n; \psi)}{\partial \psi} \right|_{\psi=\bar{\psi}} &= -\frac{1}{2} \frac{\partial}{\partial \psi} \sum_{t=1}^n \left[\log |\mathbf{H}_t| + \log |\mathbf{Q}_t| \right. \\
&\quad \left. \text{tr}[(\hat{\epsilon}_t \hat{\epsilon}_t' + \text{Var}(\epsilon_t|\mathbf{Y}_n))\mathbf{H}_t^{-1}] \right. \\
&\quad \left. \text{tr}[(\hat{\eta}_t \hat{\eta}_t' + \text{Var}(\eta_t|\mathbf{Y}_n))\mathbf{Q}_t^{-1}] \right] \Big|_{\psi=\bar{\psi}}
\end{aligned}$$

Using simple rules of matrix calculus (for more details see Magnus and Neudecker (1988)) and the results from the disturbance smoother (2.4.29)

we have,

$$\begin{aligned}
\left. \frac{\partial \log f(\mathbf{Y}_n; \boldsymbol{\psi})}{\partial \boldsymbol{\psi}} \right|_{\boldsymbol{\psi}=\tilde{\boldsymbol{\psi}}} &= -\frac{1}{2} \sum_{t=1}^n \left[\mathbf{H}_t^{-1} \frac{\partial \mathbf{H}}{\partial \boldsymbol{\psi}} - \text{tr}[\mathbf{H}_t^{-1}(\mathbf{H}_t(\mathbf{u}_t \mathbf{u}_t' - \mathbf{D}_t)\mathbf{H}_t)\mathbf{H}_t^{-1} \frac{\partial \mathbf{H}}{\partial \boldsymbol{\psi}}] \right. \\
&\quad \left. - \text{tr}[\mathbf{H}_t^{-1} \mathbf{H}_t \mathbf{H}_t^{-1} \frac{\partial \mathbf{H}}{\partial \boldsymbol{\psi}}] \right] \\
&\quad - \frac{1}{2} \sum_{t=1}^n \left[\mathbf{Q}_t^{-1} \frac{\partial \mathbf{Q}}{\partial \boldsymbol{\psi}} - \text{tr}[\mathbf{Q}_t^{-1}(\mathbf{Q}_t(\mathbf{r}_t \mathbf{r}_t' - \mathbf{N}_t)\mathbf{Q}_t)\mathbf{Q}_t^{-1} \frac{\partial \mathbf{Q}}{\partial \boldsymbol{\psi}}] \right. \\
&\quad \left. - \text{tr}[\mathbf{Q}_t^{-1} \mathbf{Q}_t \mathbf{Q}_t^{-1} \frac{\partial \mathbf{Q}}{\partial \boldsymbol{\psi}}] \right] \\
&= -\frac{1}{2} \sum_{t=1}^n \left[\mathbf{H}_t^{-1} \frac{\partial \mathbf{H}}{\partial \boldsymbol{\psi}} - \text{tr}[(\mathbf{u}_t \mathbf{u}_t' - \mathbf{D}_t) \frac{\partial \mathbf{H}}{\partial \boldsymbol{\psi}}] - \text{tr}\left[\frac{\partial \log |\mathbf{H}|}{\partial \boldsymbol{\psi}}\right] \right. \\
&\quad \left. - \frac{1}{2} \sum_{t=1}^n \left[\mathbf{Q}_t^{-1} \frac{\partial \mathbf{Q}}{\partial \boldsymbol{\psi}} - \text{tr}[(\mathbf{r}_t \mathbf{r}_t' - \mathbf{N}_t) \frac{\partial \mathbf{Q}}{\partial \boldsymbol{\psi}}] - \text{tr}\left[\frac{\partial \log |\mathbf{Q}|}{\partial \boldsymbol{\psi}}\right] \right] \right. \\
&= \frac{1}{2} \sum_{t=1}^n \text{tr} \left\{ (\mathbf{u}_t \mathbf{u}_t' - \mathbf{D}_t) \frac{\partial \mathbf{H}_t}{\partial \boldsymbol{\psi}} \right\} \\
&\quad + \frac{1}{2} \sum_{t=2}^n \text{tr} \left\{ (\mathbf{r}_{t-1} \mathbf{r}_{t-1}' - \mathbf{N}_{t-1}) \frac{\partial \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t'}{\partial \boldsymbol{\psi}} \right\} \Big|_{\boldsymbol{\psi}=\tilde{\boldsymbol{\psi}}} \quad (2.7.7)
\end{aligned}$$

The quantities \mathbf{u}_t , \mathbf{D}_t , \mathbf{r}_{t-1} , and \mathbf{N}_{t-1} are calculated during a run of the Kalman filter and smoother. The quantities $\frac{\partial \mathbf{H}_t}{\partial \boldsymbol{\psi}}$, and $\frac{\partial \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t'}{\partial \boldsymbol{\psi}}$ are usually easily calculated. In §3.2.4, we derive these quantities for periodic models we use in this thesis. We see then that calculating the score vector in state-space models is a straightforward process.

2.8 Diagnostic checking and model selection

If our model is well specified then, given the parameter values are known, the residuals are normally distributed and serially independent. Our main diag-

nostic tool is the one-step ahead prediction error which is obtained by a run of the Kalman filter. We define the standardised one-step ahead prediction error:

$$\mathbf{e}_t = F_t^{\frac{1}{2}} \mathbf{v}_t$$

which will follow a standard normal distribution. In the multivariate case it is preferred to perform the following tests for each series separately, using the standardised residual sequence of each series. Thus, without loss of generality we take e_t to be a single element of \mathbf{e}_t . We then test the normality assumption by using tests defined by Bowman and Shenton (1975). We first define the sample moments of the standardised prediction error:

$$m_1 = \frac{1}{n} \sum_{t=1}^n e_t$$

$$m_r = \frac{1}{n} \sum_{t=1}^n (e_t - m_1)^r$$

We also define the measures for skewness and kurtosis as:

$$S = \frac{m_3}{\sqrt{m_2^3}}$$

$$K = \frac{m_4}{m_2^2}$$

Then under the null hypothesis of normality $S \sim N(0, \frac{6}{n})$ and $K \sim N(3, \frac{24}{n})$.

We can also combine S and K in one statistic:

$$N = n \left\{ \frac{S^2}{6} + \frac{(K - 3)^2}{24} \right\}$$

which has a χ^2 distribution with 2 degrees of freedom.

The main tool to check the assumption of uncorrelated residuals is the sample autocorrelation function. The sample autocorrelation function at lag j , r_j ,

is defined as:

$$r_j = \frac{1}{nm_2} \sum_{t=j+1}^n (e_t - m_1)(e_{t-j} - m_1)$$

Under the null hypothesis that e_t is a white noise process, the approximate standard error for r_j is $\frac{1}{\sqrt{n}}$. A standard test statistic for serial correlation developed by (Ljung and Box 1978) is:

$$Q(k) = n(n+2) \sum_{j=1}^k \frac{r_j^2}{n-j}$$

When dealing with competing models, we may want to measure the fit of the model under consideration. Goodness of fit measures for time series models are associated with the log-likelihood, $\log L(\mathbf{Y}_n; \hat{\boldsymbol{\psi}})$. The larger the number of parameters that a model contains the larger the log-likelihood. The Akaike Information Criterion (AIC) (Akaike 1974) gives a fair comparison between models with different number of parameters by including a penalty for model order:

$$AIC = \frac{1}{n}[-2 \log L(\hat{\boldsymbol{\psi}}; \mathbf{Y}_n) + 2(q + w)]$$

where q is the dimension of the state vector $\boldsymbol{\alpha}_t$, and w is the number of estimated hyper-parameters. For a structural time series it is often the case that $w = \dim \mathbf{Q} + \dim \mathbf{H}$. In general, a model with a smaller value of AIC is preferred.

Chapter 3

Periodic variance in one season

3.1 Introduction

Series with autocovariance structure that varies with season arise in hydrology, see for example, Hipel and McLeod (1994), Troutman (1979), Pagano (1978) and Jones and Brelsford (1967). The fact that many economic time series have one season that exhibits a higher volatility than other seasons is often overlooked (Osborn and Smith 1989). This behaviour is found in monthly production series; the variability of the index of production is higher for the month with the lowest level of production. For example, the seasonal component for August in most European countries has the lowest level within a year due to summer holiday factory shut-downs. August also shows higher variability than other months. Miron (2001) shows that this is consistent with backward L-shaped marginal cost curves. Modelling this type of behaviour correctly is of importance for forecasting and seasonal adjustment.

We show that, when one month only has different variability from others there are constraints on the seasonal models we can use.

The two most common methods for seasonal adjustment, X-12-ARIMA (Findley, Monsell, Otto, Bell, and Pugh 1998) and TRAMO-SEATS (Gomez and Maravall 1996), have substantial restrictions in modelling periodic variances. In the case of X-12-ARIMA, periodic variance is dealt by applying different seasonal moving averages to each season. This ad-hoc method has proved flexible in fitting models but provides little help in detecting seasonal heteroscedasticity. There is no attempt to understand the structure and the relationship between different seasons, a common criticism for the overall philosophy of the Census X-11 and X-12 methods. TRAMO-SEATS uses an ARIMA model based decomposition of the time series and does not include modelling periodic variances.

In this chapter we develop structural models for time series in which the variance of one season differs from the others. Burridge and Wallis (1988) include periodic variances in a structural model using dummy seasonality, as in (2.3.1). However, as we show in §3.2, neither dummy seasonality nor trigonometric seasonality (Harvey 1989) are effective in modelling seasonal series with single season heteroscedasticity. We suggest two models that permit single season heteroscedasticity as a special case. In §3.2.4 and §3.2.5, some estimation and initialisation issues for these heteroscedastic models are discussed. A likelihood ratio test for seasonal heteroscedasticity is given in §3.3, while section §3.4 provides real data examples. The final section presents conclusions. Parts of this chapter are based on Tripodis and Penzer

(2006).

3.2 Models for single season heteroscedasticity

Consider a univariate seasonal time series $\{y_t : t = 1, \dots, n\}$ with seasonal period s . A structural model for $\{y_t\}$ consists of a sum of components each representing a salient feature of the series (Harvey 1989). For example, the basic structural model is:

$$y_t = \mu_t + \gamma_t + \epsilon_t$$

where μ_t is the trend, γ_t is the seasonal component and ϵ_t is the irregular (white noise) component. In the non-periodic variance case, the seasonal difference, $\{\Delta_s \gamma_t\} = \{\gamma_t - \gamma_{t-s}\}$, is a stationary process. Periodic variances can be represented by allowing $\text{Var}(\Delta_s \gamma_t)$ to depend on season. Another approach is to allow the variance of the irregular term to be periodic. The autocorrelation structure of periodic seasonal variance and periodic irregular variance models differ considerably. Below we compare the autocorrelation functions and consider the implications for single season heteroscedasticity.

Throughout we use r_t to denote the season of the t^{th} observation,

$$r_t = \begin{cases} s & t = s, 2s, \dots \\ t \pmod{s} & \text{otherwise} \end{cases}$$

We use κ_t to denote the seasonal difference of the seasonal component, $\kappa_t = \Delta_s \gamma_t$. For notational simplicity we drop the t index on r_t and κ_t when this can be done without ambiguity.

3.2.1 Periodic variance in the seasonal component

Proietti (1998) proposes a general class of models for seasonal heteroscedasticity based on the Harrison and Stevens (1976) framework described in §2.3.3. Proietti replaces the homoscedastic variance-covariance matrix of (2.3.17) with the following:

$$\begin{aligned} \text{Var}(\boldsymbol{\omega}_t) = \mathbf{V} &= [\mathbf{D} - \frac{\mathbf{1}}{\mathbf{i}_s \mathbf{D} \mathbf{i}'_s} \mathbf{D} \mathbf{i}_s \mathbf{i}'_s \mathbf{D}] = \\ &= \frac{1}{\sum_{i=1}^s \sigma_i^2} \begin{pmatrix} \sum_{i \neq 1}^s \sigma_1^2 \sigma_i^2 & -\sigma_1^2 \sigma_2^2 & \dots & -\sigma_1^2 \sigma_s^2 \\ -\sigma_1^2 \sigma_2^2 & \sum_{i \neq 2}^s \sigma_2^2 \sigma_i^2 & \dots & -\sigma_2^2 \sigma_s^2 \\ \vdots & \vdots & \ddots & \vdots \\ -\sigma_1^2 \sigma_s^2 & \dots & \dots & \sum_{i \neq s}^s \sigma_s^2 \sigma_i^2 \end{pmatrix} \end{aligned} \quad (3.2.1)$$

where $\mathbf{D} = \text{diag}\{\sigma_1^2, \dots, \sigma_s^2\}$ and $\mathbf{i}_s = (1, 1, \dots, 1)'$ is an $s \times 1$ vector. The multivariate variance-covariance matrix enforces the constraint that $S(L)\gamma_t$ is stationary. (3.2.1) implies that $\text{Cov}(\omega_t, \omega_{t-i}) = 0$ for $i \geq s$. Note also, that σ_r^2 is not the variance of the seasonal difference for season r . In fact (Tripodis and Penzer 2006),

$$\text{Var}(\kappa_t) = \text{Var}(\Delta_s \gamma_t) = s V_{rr} = s \frac{\sigma_r^2 \sum_{k \neq r}^s \sigma_k^2}{\sum_{k=1}^s \sigma_k^2} \quad (3.2.2)$$

where V_{rr} is the r^{th} diagonal element of \mathbf{V} .

We now consider single season heteroscedasticity. Suppose, without loss of generality, that the first season has a different variance from all the others.

From (3.2.2) it is clear that single season heteroscedasticity is introduced by taking $\mathbf{D} = \text{diag}\{\sigma_1^2, \sigma_2^2, \dots, \sigma_s^2\}$; setting $\sigma_1^2 \neq \sigma_2^2$ and $\sigma_2^2 = \dots = \sigma_s^2$ yields:

$$\text{Var}(\kappa_t) = \begin{cases} sV_{11} = s(s-1)\sigma_1^2\sigma_2^2/(\sigma_1^2 + (s-1)\sigma_2^2) & r_t = 1 \\ sV_{22} = s\sigma_2^2(\sigma_1^2 + (s-2)\sigma_2^2)/(\sigma_1^2 + (s-1)\sigma_2^2) & \text{otherwise} \end{cases}$$

An interesting difference arises in this case. Comparing the first with any of the other seasons gives us (Tripodis and Penzer 2006):

$$\frac{V_{11}}{V_{22}} = \frac{(s-1)}{(s-2)q+1} \quad (3.2.3)$$

where $q = \sigma_2^2/\sigma_1^2 > 0$. It is clear that the ratio of variances is a decreasing function with respect to q with a maximum of $s-1$ when q goes to zero. We conclude that, for the model defined by (2.3.16) and (3.2.1), the variance of the seasonal difference in season 1 (the distinct season) is always less than $s-1$ times higher than that of the other seasons.

The introduction of seasonal heteroscedasticity results in periodic autocorrelation in the seasonal differences. By definition, $\mathbf{x}_t = \mathbf{x}_{t-s}$, so:

$$\kappa_t = \Delta_s \gamma_t = \mathbf{x}_t' \sum_{k=1}^s \omega_{t-k}$$

From (3.2.1), and for $h = 1, \dots, s$, we get:

$$c_\kappa(r, h) = \text{Cov}(\kappa_t, \kappa_{t-h}) = -(s-h) \frac{\sigma_r^2 \sigma_{r-h}^2}{\sum_{k=1}^s \sigma_k^2} \quad \text{for } h = 1, \dots, s-1$$

where we define $\sigma_{r-h} = \sigma_{r+s-h}$ when $h \geq r$. For $h \geq s$, we have $c_\kappa(r, h) = 0$, for all r . Thus, in the homoscedastic case, $\{\Delta_s \gamma_t\}$ is a moving average of order $s-1$. In the single season heteroscedastic case, the nature of

the periodicity in the relationship between seasons is of interest. Defining $\rho_\kappa(r, h) = \text{Corr}(\kappa_t, \kappa_{t-h})$ we have, for the season with the variance that differs from the others,

$$\rho_\kappa(1, h) = -\frac{(s-h)}{s(s-1)}\sqrt{u} \quad \text{for } h = 1, \dots, s-1$$

and $\rho(1, h) = 0$ for $h \geq s$. Here $u = V_{11}/V_{rr}$, the ratio of variances defined by (3.2.3). For the remaining seasons, that is for $r = 2, \dots, s$,

$$\rho_\kappa(r, h) = \begin{cases} \rho_\kappa(1, h) & \text{if } h = r-1 \\ q\sqrt{u} \rho_\kappa(1, h) & \text{otherwise} \end{cases}$$

Thus, the lag h correlation between two standard seasons ($r \neq 1$) is a constant multiple of lag h correlation with the first season; the value of the multiplicative factor is determined by the ratios σ_2^2/σ_1^2 and $\text{Var}(\kappa_1)/\text{Var}(\kappa_2)$.

3.2.2 A comparison of periodic seasonal and periodic irregular models

An alternative approach to model seasonal heteroscedasticity is to superimpose periodic heteroscedastic measurement noise on homoscedastic seasonality. This is similar to the deseasonalised model (Hipel and McLeod 1994) used in hydrological time series where the seasonal component γ_t is deterministic and the irregular component ϵ_t has variance $\sigma_{\epsilon,r}^2$ that depends on the season r . In our model, γ_t is allowed to be stochastic with the seasonal differences having constant variance.

In order to illustrate the differences between the periodic seasonal variance

model of §3.2.1 and the periodic irregular variance model, consider a structural model with seasonal and irregular components. Define the seasonal differences as,

structural model with seasonal and irregular components. Define the seasonal differences as,

Tr

$$z_t = \Delta_s y_t = \kappa_t + \Delta_s \epsilon_t$$

zer 2006):

$$c_z(r, h) = \text{Cov}(z_t, z_{t-h}) = \begin{cases} c_\kappa(r, 0) + 2\sigma_{\epsilon, r}^2 & h = 0 \\ c_\kappa(r, h) & h = 1, \dots, s-1 \\ -\sigma_{\epsilon, r}^2 & h = s \end{cases} \quad (3.2.4)$$

with $c_z(r, h) = 0$ for $h > s$. For the periodic seasonal variance model, $\sigma_{\epsilon, r}^2 = \sigma_\epsilon^2$ for all r , so the autocovariance function is different for each season at lags $h = 0, \dots, s-1$. For the periodic irregular variance model, $c_\kappa(r, h) = c_\kappa(1, h)$ for all h , so periodicity is restricted to the variance and the lag s covariance.

The periodic behaviour of the autocovariance function indicates that the relation between the unusual season and all other seasons in the periodic irregular variance model differs from that for the periodic seasonal variance model. In the periodic irregular variance model the relation between seasons is the same within each year.

In the periodic seasonal variance model, the relation between the unusual season and all other seasons differs from the relation between any two other seasons. If increased variability in a particular season is superimposed on the series, the periodic irregular model is more appropriate. This is common in hydrological series where high variability in a particular month is usually caused by extreme weather conditions; for example, floods usually occur in

the same month but do not necessarily happen every year. On the other hand, in many economic time series high variability in one season is endogenous to the seasonal process; for example, factory owners react to lower production in one month by adapting the production in subsequent months to bring the overall output to the desired level.

By comparing the fit of the periodic irregular variance model and the periodic seasonal variance models for a given data set, we can infer whether heteroscedasticity is endogenous or superimposed on the seasonal process. In theory it is possible to identify the appropriate model for a particular series by comparing the periodicity of the sample autocovariance with the theoretical autocovariance. There are several methods for testing for periodicities in the autocorrelation function, see for example (Vecchia and Ballerini 1991) and (Hurd and Gerr 1991). However, the power of these tests is very small for samples less than 30 years so they are impractical for many economic time series. Rather than attempt to differentiate between seasonal models using the periodicities in autocovariance, we recommend post-fit diagnostics as a means of choosing an appropriate model; see practical illustrations of §3.4.

Now consider the case where the model for our series contains a trend μ_t . In this instance, the seasonal difference is given by:

$$z_t = \Delta_s y_t = \Delta_s \mu_t + \kappa_t + \Delta_s \epsilon_t$$

Suppose that μ_t follows a local level model,

$$\mu_{t+1} = \mu_t + \eta_t, \quad \{\eta_t\} \sim \text{NID}(0, \sigma_\eta^2)$$

then, using (3.2.4), we get the autocovariance function of the seasonal differences is:

$$c_z(r, h) = \text{Cov}(z_t, z_{t-h}) = \begin{cases} s\sigma_\eta^2 + c_\kappa(r, 0) + 2\sigma_{\epsilon, r}^2 & h = 0 \\ (s-h)\sigma_\eta^2 + c_\kappa(r, h) & h = 1, \dots, s-1 \\ -\sigma_{\epsilon, r}^2 & h = s \end{cases}$$

with $c_z(r, h) = 0$ for $h > s$. The difference in the two models is that the autocovariance function of the periodic seasonal variance model is periodic in all lags up to $s-1$ while the periodic irregular has periodic autocovariance at lags 0 and s only.

If μ_t follows a local linear trend model,

$$\begin{aligned} \mu_{t+1} &= \mu_t + \beta_t + \eta_t, & \{\eta_t\} &\sim \text{NID}(0, \sigma_\eta^2) \\ \beta_{t+1} &= \beta_t + \zeta_t, & \{\zeta_t\} &\sim \text{NID}(0, \sigma_\zeta^2) \end{aligned}$$

then, using (3.2.4), we get the autocovariance function of the stationary form

Δz_t :

$$c_{\Delta z}(r, h) = \begin{cases} s\sigma_\zeta^2 + 2\sigma_\eta^2 + 2c_\kappa(r, 0) - 2c_\kappa(r, 1) + 4\sigma_{\epsilon, r}^2 & h = 0 \\ (s-1)\sigma_\zeta^2 - c_\kappa(r, 0) + 2c_\kappa(r, 1) - c_\kappa(r, 2) - 2\sigma_{\epsilon, r}^2 & h = 1 \\ (s-h)\sigma_\zeta^2 - c_\kappa(r, h-1) + 2c_\kappa(r, h) - c_\kappa(r, h+1) & h = 2, \dots, s-2 \\ \sigma_\zeta^2 - c_\kappa(r, s-2) + 2c_\kappa(r, s-1) - \sigma_{\epsilon, r}^2 & h = s-1 \\ -c_\kappa(r, s-1) + \sigma_{\epsilon, r}^2 & h = s \\ -\sigma_{\epsilon, r}^2 & h = s+1 \end{cases}$$

with $c_z(r, h) = 0$ for $h > s+1$. The difference in the two models is that the autocovariance function of the periodic seasonal variance model is periodic in all lags up to s while the periodic irregular has periodic autocovariance at lags 0, 1, $s-1$, s , and $s+1$ only. In Table (3.1) we show the lags for which the autocovariance function of the stationary form for the local level and the local linear trend is periodic for the periodic seasonal model and the periodic irregular model.

Table 3.1: Lags with periodic autocorrelation in stationary form

Model for trend	Local trend	Local linear trend
Stationary form	$\Delta_s y_t$	$\Delta \Delta_s y_t$
Per. seasonal variance	$0, 1, \dots, s - 1$	$0, 1, \dots, s$
Per. irregular variance	$0, s$	$0, 1, s - 1, s, s + 1$

3.2.3 Other seasonal models

We show that dummy and trigonometric seasonal representations are not appropriate for modelling single seasonal heteroscedasticity. Consider first the dummy seasonal case. The dummy seasonality model was used by Burridge and Wallis (1988) to account for periodic variances. They propose this framework to model cases where the final estimates of the seasonal component exhibit seasonal variation. For a deterministic seasonal model, we impose the constraint that the seasonal effects sum to zero over the seasonal period. By adding a noise term $\{\omega_t\}$, we allow seasonality to evolve over time; the resulting model is referred to as the dummy seasonal representation, which was introduced in §2.3.1,

$$S(L)\gamma_t = \sum_{j=0}^{s-1} \gamma_{t-j} = \omega_t \quad (3.2.5)$$

The seasonal difference is:

$$\Delta_s \gamma_t = \Delta S(L)\gamma_t = \omega_t - \omega_{t-1} \quad (3.2.6)$$

Thus γ_t follows a seasonal ARIMA $(0, 0, 1) \times (0, 1, 0)_s$, where the moving average part is non-invertible.

In order to model seasonal heteroscedasticity we take $\omega_t \sim \text{NID}(0, \sigma_r^2)$. Using (3.2.6),

$$c_\kappa(r, h) = \text{Cov}(\kappa_t, \kappa_{t-h}) = \begin{cases} \sigma_r^2 + \sigma_{r-1}^2 & \text{for } h = 0 \\ -\sigma_{r-1}^2 & \text{for } h = 1 \\ 0 & \text{otherwise} \end{cases} \quad (3.2.7)$$

where, as before, $\sigma_0^2 = \sigma_s^2$. Now consider attempting to represent single season heteroscedasticity. Suppose, without loss of generality, that $\text{Var}(\kappa_1) = v_1$ and $\text{Var}(\kappa_2) = \dots = \text{Var}(\kappa_s) = v_2$. From (3.2.7), $v_2 = \sigma_2^2 + \sigma_1^2 = \sigma_3^2 + \sigma_2^2 = \dots = \sigma_s^2 + \sigma_{s-1}^2$ implying that $\sigma_1^2 = \sigma_{s-1}^2$ and so $v_1 = \sigma_1^2 + \sigma_s^2 = \sigma_{s-1}^2 + \sigma_s^2 = v_2$. In summary, if $s-1$ seasons have the same variance, then all seasons have the same variance. We conclude that the dummy seasonal representation cannot be used to model single season heteroscedasticity.

In the trigonometric case, introduced in §2.3.2, the seasonal effect is the combination of $[s/2]$ cycles that is $\gamma_t = \sum_{j=1}^{[s/2]} \gamma_{j,t}$ where $[s/2]$ is the integer part of $s/2$. The j^{th} cycle has frequency $\lambda_j = 2\pi j/s$ and is generated by:

$$\begin{pmatrix} \gamma_{j,t} \\ \gamma_{j,t}^* \end{pmatrix} = \begin{pmatrix} \cos \lambda_j & \sin \lambda_j \\ -\sin \lambda_j & \cos \lambda_j \end{pmatrix} \begin{pmatrix} \gamma_{j,t-1} \\ \gamma_{j,t-1}^* \end{pmatrix} + \begin{pmatrix} \omega_{j,t} \\ \omega_{j,t}^* \end{pmatrix} \quad (3.2.8)$$

where $\{\omega_{j,t}\}$ and $\{\omega_{j,t}^*\}$ are mutually independent $\text{NID}(0, \sigma_j^2)$ processes. The component $\gamma_{j,t}^*$ appear as a matter of construction.

Consider the j^{th} cycle. If we denote (3.2.8) by $\gamma_{j,t} = \mathbf{T}_j \gamma_{j,t-1} + \omega_{j,t}$, then the matrix \mathbf{T}_j has the properties $\mathbf{T}_j^s = \mathbf{I}$ and $\mathbf{T}_j^k (\mathbf{T}_j^k)' = \mathbf{I}$ where k is a positive integer. To model seasonal heteroscedasticity we allow σ_j^2 to vary with season, that is, take $\text{Var}(\omega_{j,t}) = \text{Var}(\omega_{j,t}^*) = \sigma_{j,r}^2$. The seasonal differences are given

by:

$$\kappa_{j,t} = \Delta_s \gamma_{j,t} = [\Delta_s \gamma_{j,t}]_1 = \left[\sum_{k=0}^{s-1} \mathbf{T}_j^k \omega_{j,t-k} \right]_1$$

where $[\dots]_1$ denotes the first element. The variance of the seasonal difference is then,

$$\text{Var}(\kappa_{j,t}) = \left[\sum_{k=0}^{s-1} \mathbf{T}_j^k (\mathbf{T}_j^k)' \sigma_{j,r-k}^2 \right]_{1,1} = \sum_{k=0}^{s-1} \sigma_{j,r-k}^2$$

where $[\dots]_{1,1}$ denotes the first diagonal element. Thus despite the fact that $\text{Var}(\omega_{j,t})$ depends on season, the variance of the seasonal differences is constant for the j^{th} cycle. The overall seasonal difference, $\Delta_s \gamma_t$, as a sum of $[s/2]$ homoscedastic seasonal differences, is also homoscedastic.

Using the trigonometric model, we can assign different variances to different frequencies. Modelling this type of heteroscedasticity is more appropriate in business cycle analysis rather than in seasonal analysis. Our interests in this thesis lies on whether a particular month has a different variance rather than whether a cycle with a certain periodicity has different variance to cycles in other frequencies.

We show that there is a linear relationship between the trigonometric seasonality formulated in (2.3.15), and the periodic seasonal variance model. We first show that there is a linear relationship between this trigonometric model (2.3.15) and the Harrison and Stevens model (2.3.16). As a reminder,

we rewrite these seasonal models. The trigonometric model is:

$$\begin{aligned}\gamma_t &= \mathbf{z}'_t \boldsymbol{\tau}_t & (3.2.9) \\ \mathbf{z}'_t &= [\cos \lambda_1 t, \sin \lambda_1 t, \dots, \cos \lambda_{[s/2]} t] \\ \boldsymbol{\tau}_t &= \boldsymbol{\tau}_{t-1} + \boldsymbol{\kappa}_t, \quad \boldsymbol{\kappa}_t \sim \text{NID}(\mathbf{0}_{s-1}, \mathbf{K})\end{aligned}$$

and the Harrison and Stevens (2.3.16) seasonal model is:

$$\begin{aligned}\gamma_t &= \mathbf{x}'_t \boldsymbol{\delta}_t & (3.2.10) \\ \boldsymbol{\delta}_t &= \boldsymbol{\delta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim \text{NID}(\mathbf{0}, \boldsymbol{\Omega})\end{aligned}$$

We note that there is a linear relationship between \mathbf{z}_t in (3.2.9) and \mathbf{x}_t in (3.2.10) so that,

$$\mathbf{z}_t = \mathbf{H}' \mathbf{x}_t \quad (3.2.11)$$

where $\mathbf{H}' = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_s]$ is an $(s-1) \times s$ matrix. Using (3.2.11) we can rewrite (3.2.9) as:

$$\gamma_t = \mathbf{z}'_t \boldsymbol{\tau}_t = \mathbf{x}'_t \mathbf{H} \boldsymbol{\tau}_t$$

which implies that we can rewrite the seasonal effects in (3.2.10) as $\boldsymbol{\delta}_t = \mathbf{H} \boldsymbol{\tau}_t$ which are generated as a multivariate random walk with innovation covariance,

$$\boldsymbol{\Omega} = \mathbf{H} \mathbf{K} \mathbf{H}' \quad (3.2.12)$$

We can verify using trigonometric identities that $\sum_{j=0}^{s-1} \mathbf{z}_{t-j} = \mathbf{0}$ which means that $\mathbf{i}'_s \mathbf{H} = \mathbf{0}_{s-1}$ so that $\mathbf{i}'_s \text{Var}(\boldsymbol{\omega}_t) = 0$ is enforced. In the case of the periodic seasonal model, the relationship between \mathbf{V} (3.2.1) and \mathbf{K} (3.2.12) is established by replacing $\boldsymbol{\Omega}$ with \mathbf{V} and pre and post-multiplying both sides

of (3.2.12) by \mathbf{H}' and \mathbf{H} respectively and solving for \mathbf{K} :

$$\begin{aligned}\mathbf{K} &= (\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{V}\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1} \\ &= (\mathbf{H}'\mathbf{H})^{-1}(\mathbf{H}'\mathbf{D}\mathbf{H} - \frac{1}{i_s}\mathbf{D}\mathbf{i}'_s\mathbf{H}'\mathbf{D}\mathbf{i}_s\mathbf{D}\mathbf{H})(\mathbf{H}'\mathbf{H})^{-1}\end{aligned}$$

We can then model single season heteroscedasticity using the trigonometric seasonality formulated in (2.3.15). In this case the trigonometric seasonality is equivalent to Harrison and Stevens seasonal model and therefore we will ignore it from the following analysis.

3.2.4 Score vectors for seasonal models

In this section we derive analytic derivatives for seasonal models with heteroscedasticity, which, as we saw in §2.7, are used in the estimation process. Without loss of generality we consider the following transformation of the parameters:

$$\psi_i = \frac{1}{2} \log \sigma_i^2$$

The purpose of the transformation is to ensure that $\sigma_i^2 \geq 0$. We first consider the case of the periodic irregular variance model. As before, variances of the measurement equation error differ with season. Using the same notation as in §2.7.1 we have:

$$\frac{\partial \log L(\boldsymbol{\psi}); \mathbf{Y}_n}{\partial \psi_{\epsilon, r}} = \sigma_{\epsilon, r}^2 (\mathbf{u}\mathbf{E}_r\mathbf{u}' - \Phi\mathbf{e}'_r)$$

where \mathbf{u} and Φ are $1 \times n$ vectors containing $\mathbf{u}_1, \dots, \mathbf{u}_n$ and $\mathbf{D}_1, \dots, \mathbf{D}_n$ calculated by (2.4.29) while \mathbf{e}_r is a $1 \times n$ vector with 1 in elements $r, r +$

$s, r + 2s, \dots$ and 0 otherwise, and \mathbf{E}_r is a $n \times n$ diagonal matrix with \mathbf{e}_r in its diagonal.

For the periodic seasonal variance model, we need to estimate the matrix of partial derivatives,

$$\frac{\partial \mathbf{V}}{\partial \sigma_i^2} = \begin{pmatrix} \frac{\partial \mathbf{V}_{11}}{\partial \sigma_i} & \cdots & \frac{\partial \mathbf{V}_{1s}}{\partial \sigma_i} \\ \vdots & \ddots & \vdots \\ \frac{\partial \mathbf{V}_{s1}}{\partial \sigma_i} & \cdots & \frac{\partial \mathbf{V}_{ss}}{\partial \sigma_i} \end{pmatrix} \quad (3.2.13)$$

where $\mathbf{V} = [\mathbf{D} - \frac{1}{\mathbf{i}_s \mathbf{D} \mathbf{i}'_s} \mathbf{D} \mathbf{i}_s \mathbf{i}'_s \mathbf{D}]$ is the variance-covariance matrix of the periodic seasonal model (3.2.1). Let the function $\sigma^2(i)$ equal the i^{th} element of the diagonal of \mathbf{D} , that is $\sigma^2(i) = \mathbf{D}_{ii}$. Let \mathbf{D} have j different elements on the diagonal. Each σ_i^2 is repeated r_i times, that is if all elements of \mathbf{D} are equal except for rows k and $k-1$ where $\sigma^2(k) = \sigma^2(k-1) = \sigma_k^2$, then $r_k = 2$. Without loss of generality we assume that $\sigma^2(k) = \sigma_k^2$. From (3.2.1) we get:

$$\begin{aligned} \mathbf{V}_{kk} &= \frac{(r_k - 1)\sigma_k^4 + \sigma_k^2 \sum_{i \neq k}^j r_i \sigma_i^2}{\sum_{i=1}^j r_i \sigma_i^2} \\ \mathbf{V}_{km} &= -\frac{\sigma_k^4}{\sum_{i=1}^j r_i \sigma_i^2} \quad \text{if } \sigma^2(k) = \sigma^2(m) \\ \mathbf{V}_{km} &= -\frac{\sigma_k^2 \sigma_m^2}{\sum_{i=1}^j r_i \sigma_i^2} \quad \text{if } \sigma^2(k) \neq \sigma^2(m) \end{aligned}$$

for $k, m = 1, \dots, s$.

Using this representation and applying simple calculus we get the following

partial derivatives:

$$\frac{\partial \mathbf{V}_{kk}}{\partial \sigma^2(k)} = \frac{r_k(r_k - 1)\sigma_k^4 + \sum_{i \neq k}^j r_i \sigma_i^2 [(r_k - 2)\sigma_k^2 + \sum_{i=1}^j r_i \sigma_i^2]}{(\sum_{i=1}^j r_i \sigma_i^2)^2} \quad (3.2.14)$$

$$\frac{\partial \mathbf{V}_{kk}}{\partial \sigma^2(l)} = \frac{r_l \sigma_k^4}{(\sum_{i=1}^j r_i \sigma_i^2)^2} \quad \text{for } \sigma^2(l) \neq \sigma^2(k) \quad (3.2.15)$$

$$\frac{\partial \mathbf{V}_{km}}{\partial \sigma^2(k)} = -\frac{r_k \sigma_k^4 + 2\sigma_k^2 \sum_{i \neq k}^j r_i \sigma_i^2}{(\sum_{i=1}^j r_i \sigma_i^2)^2} \quad \text{for } \sigma^2(k) = \sigma^2(m) \quad (3.2.16)$$

$$\frac{\partial \mathbf{V}_{km}}{\partial \sigma^2(k)} = -\frac{\sigma_m^2 \sum_{i \neq k}^j r_i \sigma_i^2}{(\sum_{i=1}^j r_i \sigma_i^2)^2} \quad \text{for } \sigma^2(k) \neq \sigma^2(m) \quad (3.2.17)$$

$$\frac{\partial \mathbf{V}_{km}}{\partial \sigma^2(l)} = \frac{r_l \sigma_k^4}{(\sum_{i=1}^j r_i \sigma_i^2)^2} \quad \text{for } \sigma^2(k) = \sigma^2(m) \neq \sigma^2(l) \quad (3.2.18)$$

$$\frac{\partial \mathbf{V}_{km}}{\partial \sigma^2(l)} = \frac{r_l \sigma_k^2 \sigma_m^2}{(\sum_{i=1}^j r_i \sigma_i^2)^2} \quad \text{for } \sigma^2(k) \neq \sigma^2(m), l \neq k \text{ or } m \quad (3.2.19)$$

for $k, l, m = 1, \dots, s$.

Using (3.2.15-3.2.19), we can calculate the set of analytic derivatives in (3.2.13). This is a sub-matrix of $\frac{\partial \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t'}{\partial \boldsymbol{\psi}}$, which is used to calculate the score vector in (2.7.7), as shown in §2.7.2.

For the case of single season heteroscedasticity we have

$$\mathbf{V} = \frac{1}{\sigma_1^2 + (s-1)\sigma_2^2} \begin{pmatrix} (s-1)\sigma_1^2\sigma_2^2 & -\sigma_1^2\sigma_2^2 & \dots & -\sigma_1^2\sigma_2^2 \\ -\sigma_1^2\sigma_2^2 & (s-2)\sigma_2^2 + \sigma_1^2\sigma_2^2 & \dots & -\sigma_2^4 \\ \vdots & \vdots & \ddots & \vdots \\ -\sigma_1^2\sigma_2^2 & -\sigma_2^4 & \dots & (s-2)\sigma_2^2 + \sigma_1^2\sigma_2^2 \end{pmatrix}$$

We use (3.2.14) to calculate $\frac{\partial \mathbf{V}_{11}}{\partial \sigma_1^2}$, and $\frac{\partial \mathbf{V}_{kk}}{\partial \sigma_2^2}$ for $k \geq 2$, (3.2.15) to calculate $\frac{\partial \mathbf{V}_{11}}{\partial \sigma_2^2}$, and $\frac{\partial \mathbf{V}_{kk}}{\partial \sigma_1^2}$ for $k \geq 2$, (3.2.16) to calculate $\frac{\partial \mathbf{V}_{km}}{\partial \sigma_2^2}$ for $k, m \geq 2$, (3.2.17) to calculate $\frac{\partial \mathbf{V}_{1m}}{\partial \sigma_1^2}$, and $\frac{\partial \mathbf{V}_{1m}}{\partial \sigma_2^2}$ for $m \geq 2$, (3.2.18) to calculate $\frac{\partial \mathbf{V}_{km}}{\partial \sigma_1^2}$ for $k, m \geq 2$.

We then have

$$\frac{\partial \mathbf{V}}{\partial \sigma_1^2} = \frac{\sigma_2^4}{(\sigma_1^2 + (s-1)\sigma_2^2)^2} \begin{pmatrix} (s-1)^2 & -(s-1) & \dots & -(s-1) \\ -(s-1) & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ -(s-1) & 1 & \dots & 1 \end{pmatrix}$$

and

$$\frac{\partial \mathbf{V}}{\partial \sigma_2^2} = \frac{\sigma_1^2\sigma_2^2}{(\sigma_1^2 + (s-1)\sigma_2^2)^2} \begin{pmatrix} (s-1)q^{-1} & -q^{-1} & \dots & -q^{-1} \\ -q^{-1} & \frac{\partial \mathbf{V}_{22}}{\partial \sigma_2} & \dots & -(s-1)q - 2 \\ \vdots & \vdots & \ddots & \vdots \\ -q^{-1} & -(s-1)q - 2 & \dots & \frac{\partial \mathbf{V}_{22}}{\partial \sigma_2} \end{pmatrix}$$

with $\frac{\partial \mathbf{V}_{22}}{\partial \sigma_2} = (s-1)(s-2)q + 2(s-2) + q^{-1}$ and $q = \sigma_2^2/\sigma_1^2$

3.2.5 Initialisation of HS

For HS seasonality we need to ensure that the block in $\mathbf{P}_{\infty,1}$, from (2.6.1), that relates to the seasonal component, which we denote $\mathbf{P}_{[\gamma]}$, is a symmetric matrix of the form (2.3.17). For example:

$$\mathbf{P}_{[\gamma]} = [\mathbf{I}_s - \frac{1}{s}\mathbf{i}_s\mathbf{i}_s'] = \begin{pmatrix} 1 - \frac{1}{s} & -\frac{1}{s} & \dots & -\frac{1}{s} \\ -\frac{1}{s} & 1 - \frac{1}{s} & \dots & -\frac{1}{s} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{s} & \dots & \dots & 1 - \frac{1}{s} \end{pmatrix}$$

This ensures that the state variance $\mathbf{P}_{[\gamma]}$ of the multivariate random walk in the seasonal component is not of full rank, as required by (3.2.1).

3.3 Test for seasonal heteroscedasticity

Despite the fact that seasonal heteroscedasticity is relatively common, there are few methods to test for its presence. Existing tests are based on the likelihood ratio, Wald or Lagrange multiplier principles (Engle 1984). In practice, some version of Goldfeld and Quandt (1965) or White (1980) test for heteroscedasticity, adjusted for seasonal series is used; these are misspecification tests rather than tests of a specific hypothesis. Useful information about seasonal heteroscedasticity can also be obtained graphically from inspection of time series plots, correlograms, periodograms of the squared data and seasonal sub-plots. In figure (3.1), we show all different types of graphs

for the index of production for Italy. We see from the periodogram of the squared data that there is a significant peak at the seasonal frequency which indicates seasonal heteroscedasticity. From the plot of the original series and the seasonal subplot, we see that this is concentrated in the month of August for reasons we examine in the next section. In this section we suggest a likelihood ratio test for testing the hypothesis that one month exhibits a different variance than the others, under the null hypothesis that all seasons have the same variance. With a parameter vector $\boldsymbol{\psi}$, we denote the likelihood function of the null model as L_0 while the likelihood function of the alternative model that one month has a different variance we denote as L_1 . The likelihood ratio test statistics is then (Hamilton 1994, p.144):

$$LR = 2(\log L_1 - \log L_0)$$

which is asymptotically distributed as χ_1^2 under the null hypothesis. We use a Monte Carlo experiment to find the approximate power of our test. In each experiment 10000 replications of a basic structural model with local level trend component and quarterly Harrison and Stevens (3.2.10) seasonal component are simulated (Tripodis and Penzer 2006). We also did some simulations for monthly data, but since the results were not affected by the periodicity, we only present the results for the quarterly data. For a given simulation, the null is rejected if the test statistic exceeds the 95th percentile of a χ_1^2 distribution. We evaluate the rejection frequency for each combination of sample size and parameter value. Table 3.2 shows the results for the periodic irregular model in which the irregular component has variance $\sigma_{\epsilon,1}^2$ in season 1 and $\sigma_{\epsilon,2}^2$ otherwise. Table 3.3 refers to the periodic seasonal variance

Figure 3.1: Graphic diagnostics for seasonal heteroscedasticity (index of production for Italy)

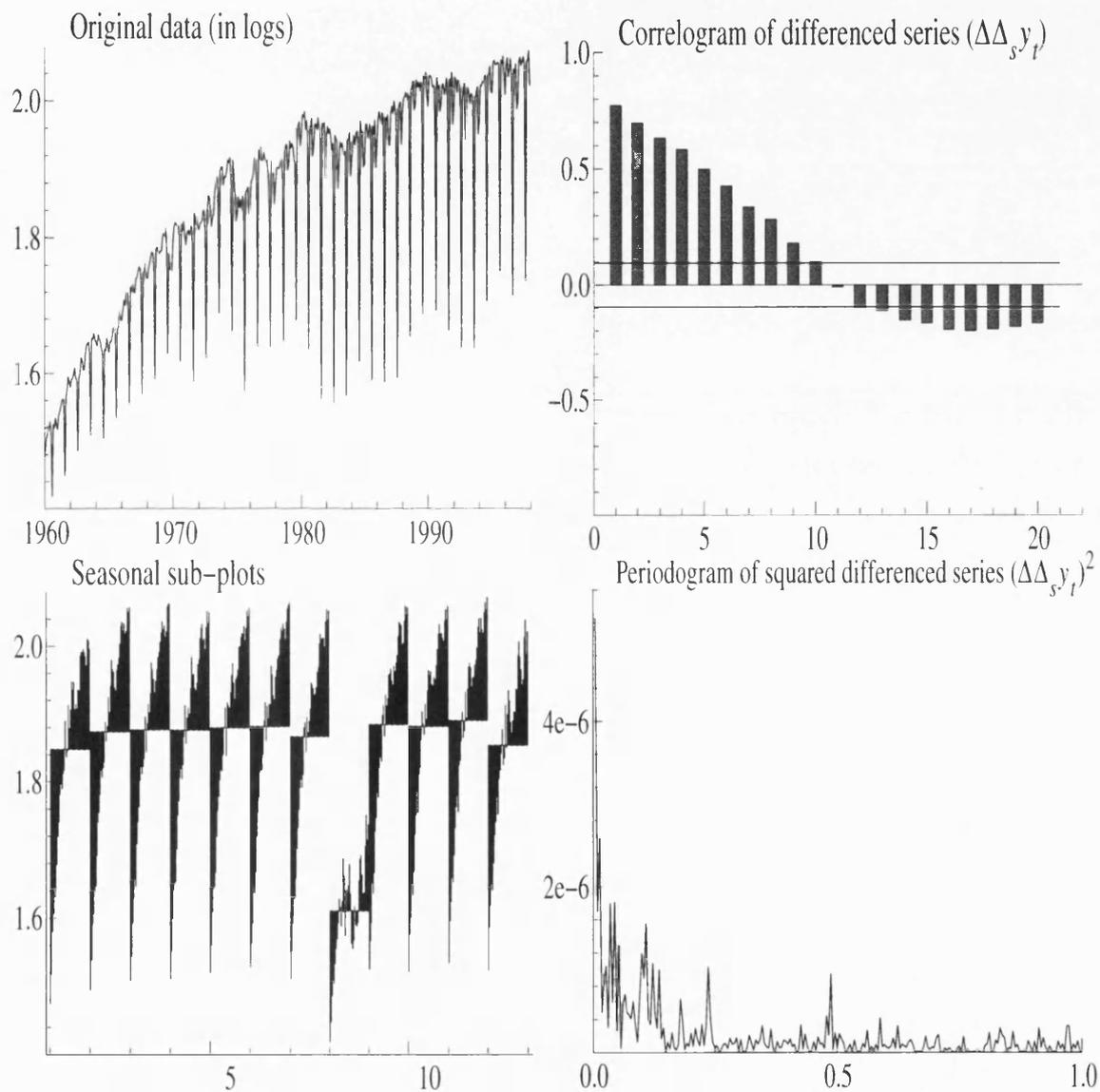


Table 3.2: Rejection frequency (%) for periodic irregular variance model

$\sigma_{\epsilon,1}^2$	N=40	N=80	N=120	N=240	N=480
0.1	9.67	23.3	38.6	69.9	95.0
0.5	5.14	9.41	13.3	22.2	39.0
1	4.03	5.03	4.68	4.49	5.00
2	6.54	9.17	12.7	40.9	68.8
5	44.9	74.4	89.2	99.5	100
10	79.5	97.5	99.8	100	100
100	99.8	100	100	100	100

10000 replications of a local level model (2.3.1)+ Harrison and Stevens (3.2.10) seasonality with periodic irregular variance. $\sigma_1^2 = 0.1$, $\sigma_{\epsilon,2}^2 = 1$ and $\sigma_\eta^2 = 1$

case with the seasonal variance in the first season, σ_1^2 , different from the other season's variance, σ_2^2 . Note that, in the periodic seasonal variance case, we take $\sigma_{\epsilon,2}^2$ to be the variance of the irregular term and, in the periodic irregular variance case, we take σ_2^2 to be the variance of the noise term associated with the seasonal component.

The rejection frequency when the null hypothesis is true indicates that χ_1^2 provides a reasonable approximation to the distribution of the test statistic. The power of the test appears reasonable except in two instances. In the periodic irregular variance case the test does not perform well when the variance of the unusual season is smaller than the other variances. For the periodic seasonal variance model, the power is low for sample sizes less than 200.

Table 3.3: Rejection frequency (%) for periodic seasonal variance model

σ_1^2	N=40	N=80	N=120	N=240	N=480
0.1	29.1	64.4	83.9	99.3	100
0.5	9.06	14.7	18.2	29.6	54.1
1	6.2	6.9	5.8	5.6	5.05
2	8.61	14.1	17.4	28.7	50.9
5	18.3	37.1	51.5	80.7	97.9
10	25.1	51.4	69.2	93.8	99.8
100	35.9	68.5	84.7	98.8	100

10000 replications of a local level model (2.3.1)+ Harrison and Stevens (3.2.10) seasonality with periodic seasonal variance. $\sigma_\epsilon^2 = 1$, $\sigma_2^2 = 1$, and $\sigma_\eta^2 = 1$

Table 3.4: Rejection frequency (%) for periodic irregular variance model : comparing across signal to noise ratios

$\sigma_{\epsilon,1}^2$	$q_1 = 1, q_2 = 0.1$		$q_1 = 0.1, q_2 = 1$	
	N=120	N=240	N=120	N=240
0.1	1.77	2.96	38.7	71.2
0.5	22.9	81.5	86.9	99.7
1	63.6	87.2	96.1	99.9
5	81.4	97.1	98.7	100

10000 replications of a local level model (2.3.1)+ Harrison and Stevens (3.2.10) seasonality with periodic irregular variance and $\sigma_\eta^2 = 1$

Table 3.5: Rejection frequency (%) for periodic seasonal variance model : comparing across signal to noise ratios

σ_1^2	$q_1 = 1, q_2 = 0.1$		$q_1 = 0.1, q_2 = 1$	
	N=120	N=240	N=120	N=240
0.1	53.4	86.4	2.63	5.66
0.5	89.4	99.7	4.71	18.6
1	99.9	100	10.5	28.8
5	99.9	100	31.9	59.7

10000 replications of a local level model (2.3.1)+
Harrison and Stevens (3.2.10) seasonality with
periodic seasonal variance and $\sigma_\eta^2 = 1$

The results depend in part on the contribution of the seasonal variation to the total variation. In order to illustrate sensitivity to changes in the signal-to-noise ratios, we run two further experiments. Define the following signal-to-noise ratios: $q_1 = \sigma_2^2/\sigma_{\epsilon,1}^2$ and $q_2 = \sigma_2^2/\sigma_{\epsilon,2}^2$ for the periodic irregular variance model; $q_1 = \sigma_1^2/\sigma_{\epsilon,2}^2$ and $q_2 = \sigma_2^2/\sigma_{\epsilon,2}^2$ for the periodic seasonal variance model. We consider two pairings of the signal-to-noise ratio: $q_1 = 1, q_2 = 0.1$ and $q_1 = 0.1, q_2 = 1$. In the periodic irregular case, varying $\sigma_{\epsilon,1}^2$ allows us to investigate the effect of altering the values of the other parameter value since $\sigma_2^2 = q_1\sigma_{\epsilon,1}^2$ and $\sigma_{\epsilon,2}^2 = q_1\sigma_{\epsilon,1}^2/q_2$. In the periodic seasonal variance case we vary σ_1^2 and use the relationships $\sigma_{\epsilon,2}^2 = \sigma_1^2/q_1$ and $\sigma_2^2 = q_2\sigma_1^2/q_1$. Tables 3.4 and 3.5 show the rejection frequency (%) across signal-to-noise ratios in 10000 replications of a model with local level (2.3.1) and quarterly Harrison

and Stevens seasonality (3.2.10). The variance of the level component is held fixed. All tables in this section are from Tripodis and Penzer (2006).

Table 3.4 shows that the rejection frequency for the periodic irregular variance case is generally high irrespective of the values of the hyperparameters. There are two exceptions. When $\sigma_{\epsilon,1}^2$ is small the test tends to perform badly. Also in the case where the sample size is relatively small and q_1 is large relative to q_2 , so the variance of the unusual season is smaller than the variance of the other seasons, the rejection frequency is strongly dependent on the value of $\sigma_{\epsilon,1}^2$. Table 3.5 shows that similar results hold for the periodic seasonal variance model. In this case, when q_2 is larger than q_1 the variance associated with the unusual season is small relative to the variance of the other seasons.

3.4 Applications

To illustrate our approach to single season heteroscedasticity we consider the monthly index of production for Italy, France and Spain. Figure 3.2 shows the monthly Index of Production for Italy from January 1960 to December 1997 and for France and Spain from January 1960 to January 2003. As mentioned before, economic theory predicts that the lowest season will exhibit higher variability. In all three countries production is lowest in August which coincides with the holiday season. We also showed in §3.3 that we can use graphs to identify the season which has a different variance for all other

periods. From figure (3.1), we saw that August is the month with a different variance for the monthly Index of Production for Italy. We get similar graphs for all the other countries' indices. We fit a basic structural model with local linear trend (2.3.1) and Harrison and Stevens seasonality (3.2.10). Three cases are considered, homoscedastic, periodic irregular variance and periodic seasonal variance. Signal extraction is performed using the Kalman filter and is implemented in Ox (Doornik 1998). Maximum likelihood estimates of the parameters for Italy, France are presented in Tables 3.6, 3.7 and 3.8 respectively. In these tables, $\sigma_{\epsilon,1}^2$ and σ_1^2 are respectively the irregular variance and the seasonal variance for August. The likelihood ratio test for both heteroscedastic models is significant for Italy and France while for Spain only the periodic seasonal variance model is significant. We note that the critical value for $\alpha = 0.05$ is $\chi_1^2 = 3.84$. According to the AIC criterion, periodic seasonal variance is preferred for all three series. Even though the ratio of the seasonal variances is very high, it does not create a problem for distinguishing heteroscedastic from homoscedastic cases in the HS model. Figures (3.3)-(3.5) show some residual diagnostic graphs for the series considered using the model with the best fit (in the case, the periodic seasonal model). We see that there is no residual pattern in the correlogram or the periodogram of the residuals, while the histogram show that the residuals are roughly normal. All these diagnostic graphs indicate that the model is satisfactory. The better fit of the periodic seasonal variance model indicates that high variability in August is a feature of the seasonal component which is balanced across all seasons.

Table 3.6: Italian Index of Production: parameter estimates and diagnostic statistics

	Homoscedastic	Periodic irregular	Periodic seasonal
$\sigma_{\epsilon,1}^2$		0.007	
$\sigma_{\epsilon,2}^2$	3.0×10^{-5}	3.7×10^{-5}	2.1×10^{-5}
σ_{η}^2	4.7×10^{-5}	5.0×10^{-5}	4.6×10^{-5}
σ_{ζ}^2	4.9×10^{-9}	1.0×10^{-10}	2.7×10^{-13}
σ_1^2			4.3×10^{-5}
σ_2^2	6.2×10^{-6}	6.1×10^{-7}	1.6×10^{-6}
LR		75	114
r_1^a	-0.002	0.007	0.026
r_{12}^b	0.03	0.08	0.02
$Q(12)^c$	15.3	21** ^g	15
S^d	-0.8	0.0	-0.4
K^e	7.7**	4.2**	5.1**
N^f	450**	25**	97**
AIC	-13.9	-14.3	-14.9

^a Residual autocorrelation at lag 1 (see §2.8)

^b Residual autocorrelation at lag 12 (see §2.8)

^c Ljung-Box statistic based on 12 residual autocorrelations. It is asymptotically χ^2 with degrees of freedom given by $12 - n^*$ where n^* is the number of hyperparameters excluding $\sigma_{\epsilon,2}^2$ (Harvey 1989, p.259)

^d Test for residual skewness (see §2.8)

^e Test for residual kurtosis (see §2.8)

^f Bowman-Shenton test for non-normality (see §2.8)

^g Significant for $\alpha = 0.05$

Table 3.7: French Index of Production: parameter estimates and diagnostic statistics

	Homoscedastic	Periodic irregular	Periodic seasonal
$\sigma_{\epsilon,1}^2$		0.001	
$\sigma_{\epsilon,2}^2$	0.0001	0.0002	0.0002
σ_{η}^2	7.9×10^{-5}	7.2×10^{-5}	6.8×10^{-5}
σ_{ζ}^2	3.4×10^{-8}	3.8×10^{-8}	4.1×10^{-8}
σ_1^2			0.0001
σ_2^2	2.3×10^{-5}	1.6×10^{-5}	6.1×10^{-6}
LR		17	49
r_1	-0.10**	-0.08	-0.08
r_{12}	0.06	0.10**	0.15**
Q(12)	104**	118**	143**
S	-0.3**	-0.2	-0.2
K	4.3**	3.6**	3.6**
N	42.3**	11.5**	10.4**
AIC	-11.9	-12	-12.1

A correction for outliers in 1968.5, 1968.6 and a level shift in 1974.11 are included

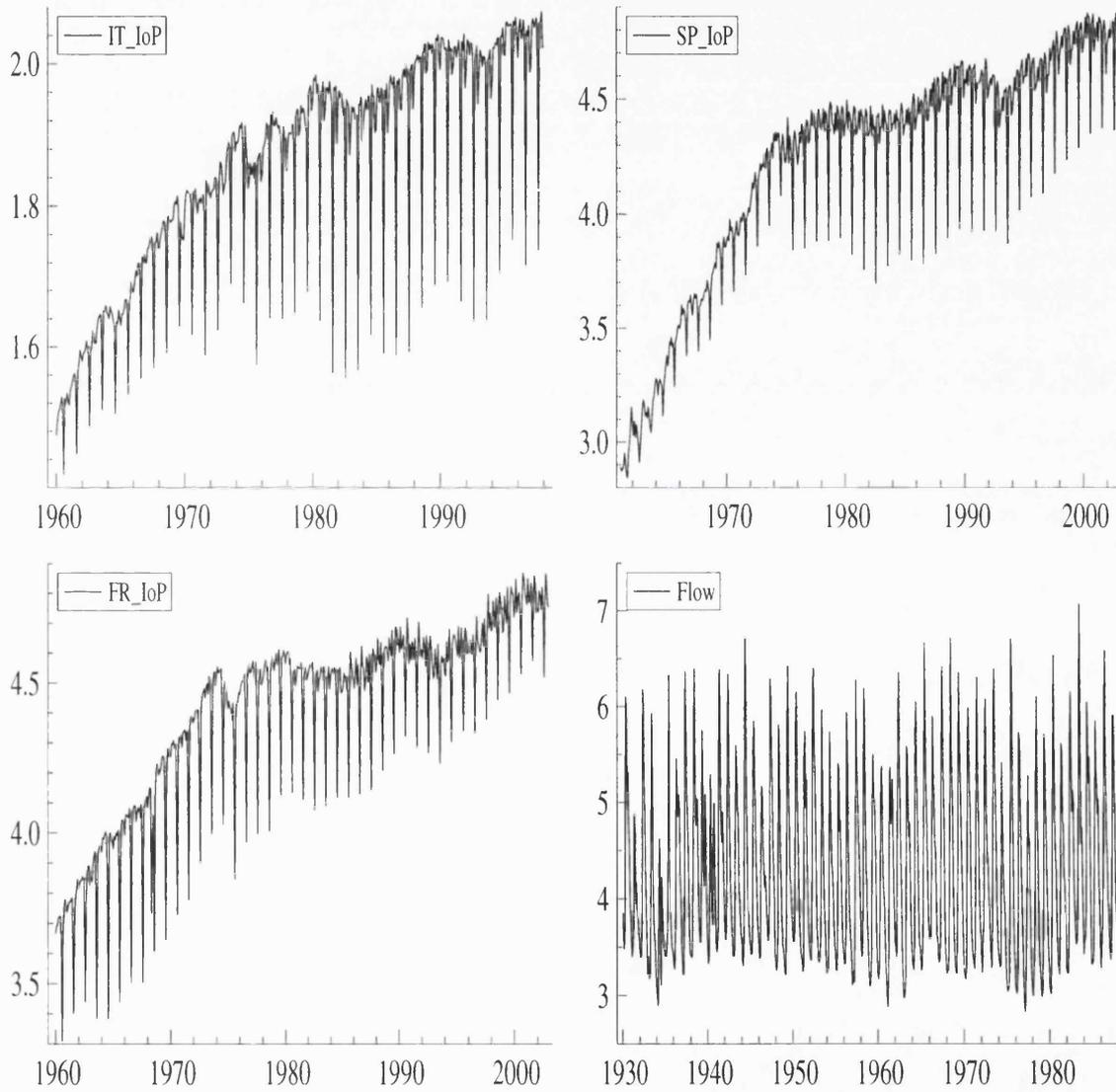
Table 3.8: Spanish Index of Production: parameter estimates and diagnostic statistics

	Homoscedastic	Periodic irregular	Periodic seasonal
$\sigma_{\epsilon,1}^2$		0.016	
$\sigma_{\epsilon,2}^2$	0.0002	0.0002	0.0002
σ_{η}^2	0.0001	0.0001	0.0001
σ_{ζ}^2	1.6×10^{-7}	1.5×10^{-7}	1.6×10^{-7}
σ_1^2			0.0002
σ_2^2	3.0×10^{-5}	1.3×10^{-5}	1.1×10^{-6}
LR		0	77
r_1	-0.05	-0.04	-0.02
r_{12}	0.004	0.06	0.03
Q(12)	15.7	6.7	7.7
S	-0.6**	-0.1	-0.3**
K	6.2**	3.9**	4.7**
N	240**	18**	64**
AIC	-11.1	-11.1	-11.4

Table 3.9: River flow in Whiterocks river: parameter estimates and diagnostic statistics

	Homoscedastic	Periodic irregular	Periodic seasonal
$\sigma_{\epsilon,1}^2$		0.22	
$\sigma_{\epsilon,2}^2$	0.05	0.03	0.15
σ_{η}^2	0.03	0.03	0.03
σ_1^2			0.05
σ_2^2	3.3×10^{-5}	3.0×10^{-5}	2.5×10^{-5}
LR		74.1	1.13
r_1	0.17**	0.17**	0.18**
r_{12}	-0.14**	-0.10**	-0.14**
Q(12)	96.3**	107.0**	101**
S	0.6**	0.6**	0.5**
K	5.9	5.1	5.7
N	298**	168**	234**
AIC	-2.2	-2.4	-2.2

Figure 3.2: Time series used in applications section : index of production for Italy, France and Spain, and Whiterocks river flow



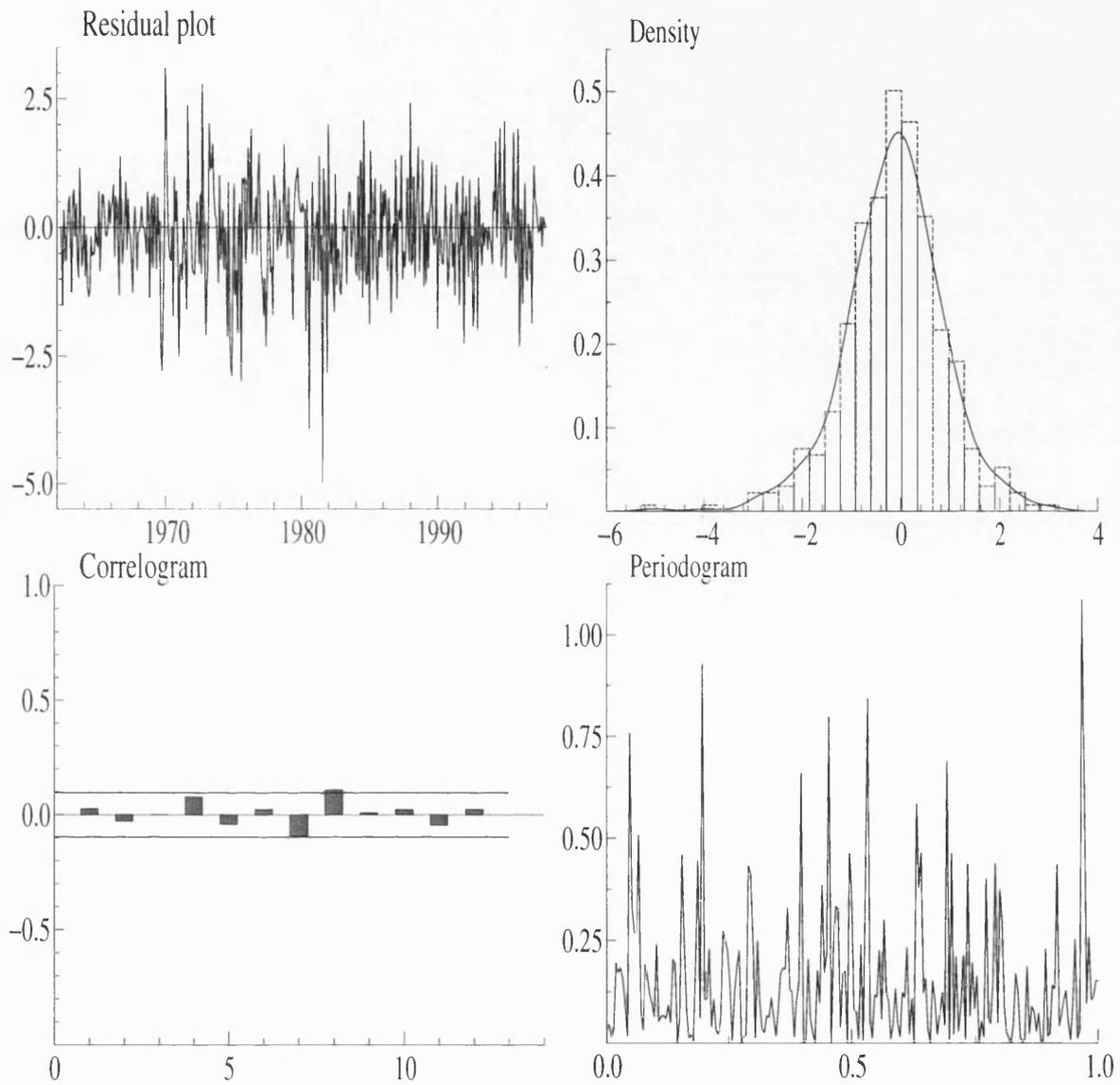


Figure 3.3: Residuals diagnostics for index of production for Italy

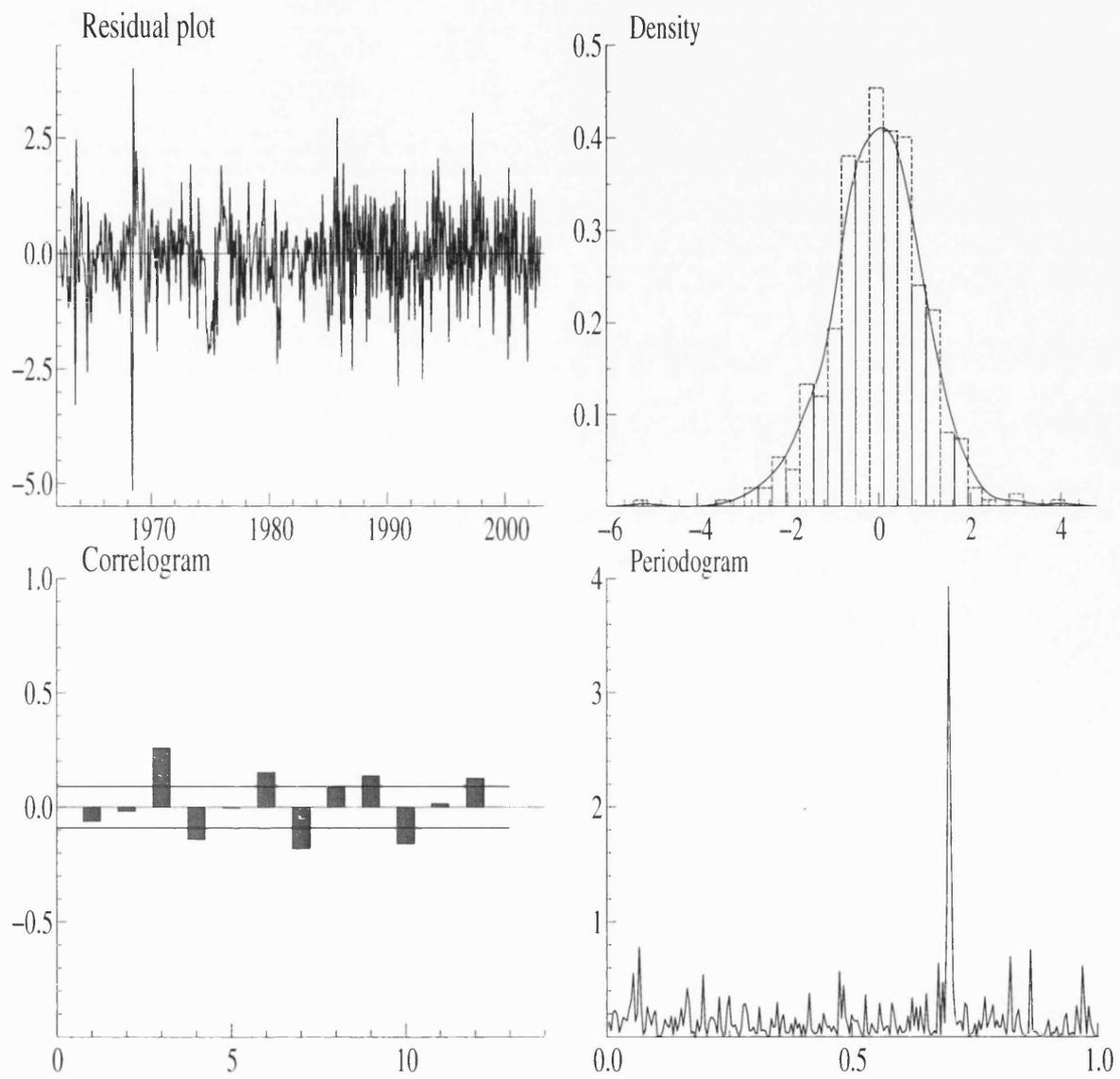


Figure 3.4: Residuals diagnostics for index of production for France

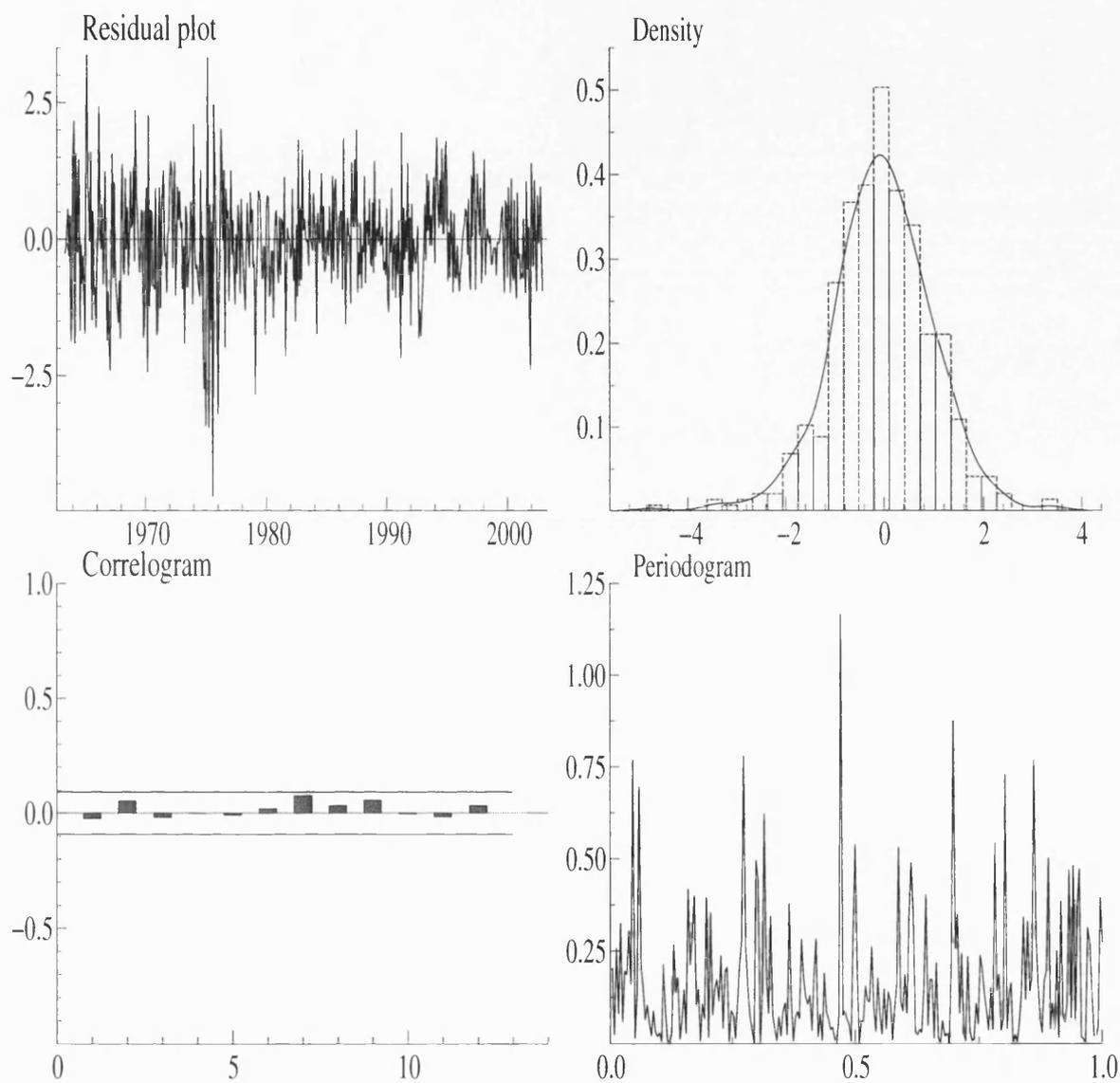


Figure 3.5: Residuals diagnostics for index of production for Spain

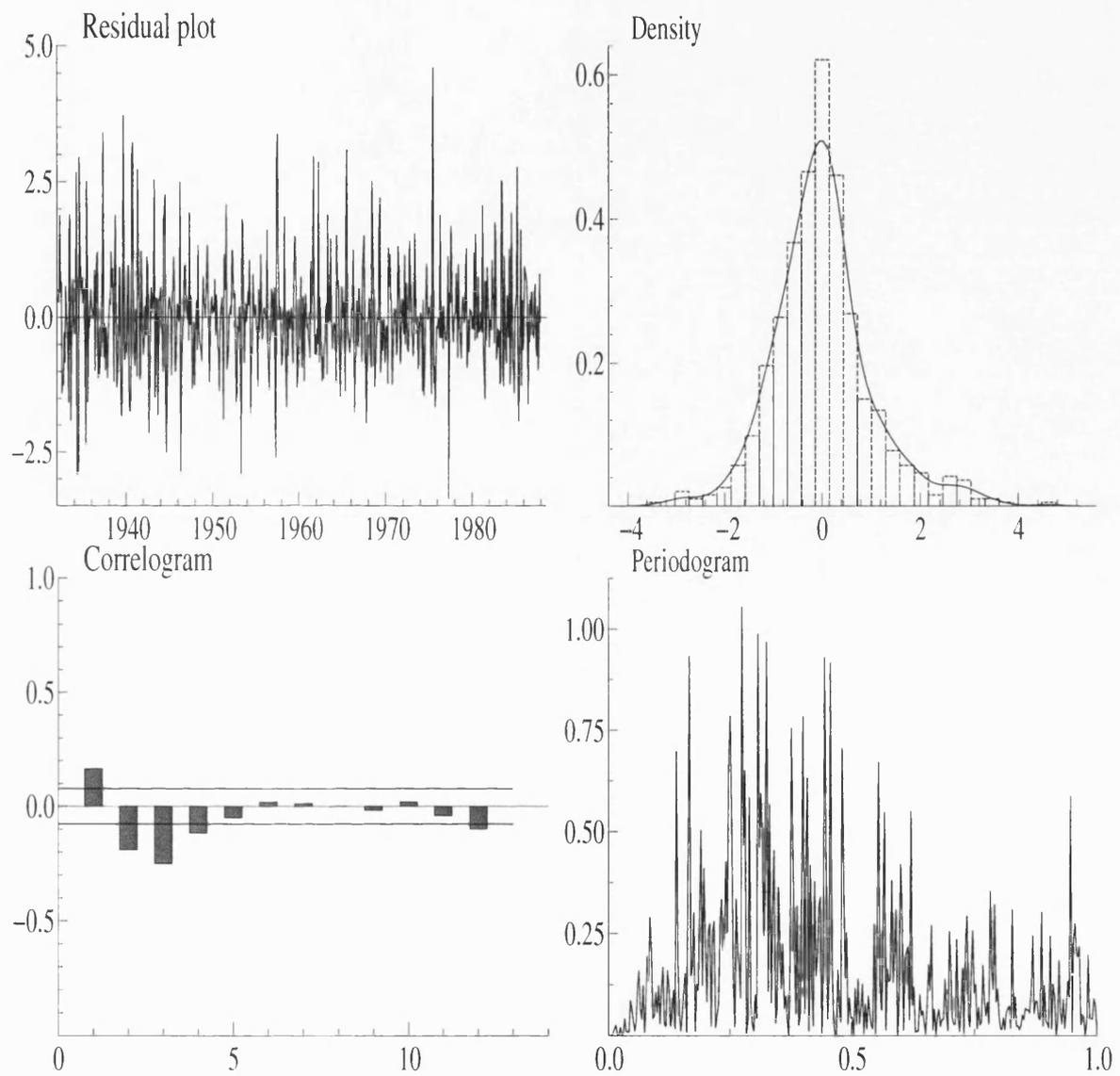


Figure 3.6: Residuals diagnostics for Whiterocks river flow

We now turn to a case where the periodic irregular variance model is more appropriate. Figure 3.2 displays the monthly river flow in feet in logarithms for the Whiterocks river (Utah). The peak in river flow occurs in June. A local level provides the best model for the trend in this example. Parameter estimates from fitting homoscedastic, periodic irregular variance and periodic seasonal variance models are given in table 3.9. The null hypothesis of homoscedasticity is rejected when the alternative is a periodic irregular variance model but not when the alternative is a periodic seasonal variance model. This indicates that the higher variability in one month is superimposed in the series. The periodic irregular variance also has better fit, although in this instance neither model fits particularly well. Figure (3.6 show diagnostic graphs similar to the ones examined before. Closer inspection of the correlogram and the periodogram of the residuals, show that there is probably a cyclical (non-seasonal) component in the series. Cyclical (non-seasonal) components are beyond the scope of this thesis and they were not examined but the inclusion of such a component would probably improve considerably the diagnostics of this series.

3.5 Conclusions

Economic time series often have one season with different variance from the others. Reasonable models of seasonal variability yield useful descriptive information and provide a basis for season dependent prediction interval estimation. Single season heteroscedasticity can be represented using peri-

odic seasonal variance or periodic irregular variance; these approaches differ markedly in their periodic covariance structure. We suggest that periodic seasonal variance is more appropriate for modelling economic time series. In economic series, higher variability in one season is usually a feature of the seasonal component; economic agents have knowledge of seasonality and are able to counteract higher variability in one season by adjusting their behaviour in all other seasons. For example, factory owners may compensate for low production in one month by increased production in subsequent months. Higher variability in a single season may also result from an exogenous effect. In the case of river flow, the exogenous effect is rainfall which is not balanced across the year. For series where there is no mechanism to balance higher variability in one season across the other seasons, a periodic irregular model may be more appropriate. We show that a likelihood ratio test is effective in detecting single season heteroscedasticity. Our test also allows us to distinguish between cases with periodic variance in the irregular component and those where periodic seasonal variance provides a better model. For practical purposes, we suggest that both the periodic seasonal and the periodic irregular models are fit in the data and an information criterion, such as AIC, is used to determine the appropriate model. If the season with the different variance is not known beforehand, then graphical techniques, such as seasonal sub-plots and periodograms can be used to determine the nature of the seasonal heteroscedasticity.

Chapter 4

Structural time series models for periodic processes

4.1 Introduction

Economists traditionally view seasonality as a redundant feature of a time series that needs to be removed before economic analysis (Osborn and Smith 1988). In recent years there has been increased interest in modelling seasonality and an understanding that economic analysis could be flawed if seasonality is ignored, (Hylleberg 1992; Ghysels and Osborn 2001) and references therein. Seasonality is usually viewed as an unobserved component with constant variance and zero sum over the seasonal period (Bell and Hillmer 1984). Our interest lies in seasonal time series with periodicity in the second moments, that is in the autocovariance function. This gives rise to models for which the confidence interval of the forecasts is season dependent. Models with a periodic autocovariance function have been investigated within the

autoregressive moving average (ARMA) framework (Pagano 1978; Parzen and Pagano 1979). Periodic ARMA are specified in a similar way to non-periodic ARMA models but the former have parameters which change with season. Most applications exclude the MA part for ease of estimation. Periodic AR models (PAR) have been used successfully in economic time series. Novales and de Frutto (1997), Franses (1996), Osborn and Smith (1989) show that a large proportion of macroeconomic time series have periodic second moments.

An alternative class of time series models are the structural time series models (Harvey 1989). Following this methodology, the salient features of the data such as trend, seasonal and irregular are modelled directly as stochastic processes. The model is cast in state-space form, and the Kalman filter is used for estimation. Structural time series models (STSM) can be extended to incorporate seasonality in the second moments. We propose a new class of periodic structural time series models (PSTSM) and show that PSTSM are observationally equivalent to periodic integrated moving average (PIMA) models.

Periodic models are efficiently represented in a vector form with the time index measured in years and estimated using multivariate analysis (Gladyshchev 1961). PAR models and their vector representation are described in §4.2 along with PMA models. §4.3 describes the extensions of STSM to the periodic case and the relation between PSTSM and PARMA models. §4.4 looks at the forecasting accuracy of PAR and PSTSM. We compare PAR and PSTSM on a data set of eleven quarterly macroeconomic variables from

USA, Canada, Germany, and UK. PSTSM produce better forecasts, both within and out of sample, for the majority of the series concerned. The final section presents conclusions.

4.2 Periodic AR and MA models

Consider an observed time series $y_{s,n}$, where $s = 1, \dots, S$ denotes the season and $n = 1, \dots, N$ the year. A simple periodic AR(1), or PAR(1), has the form:

$$y_{s,n} = \phi_{1,s}y_{s-1,n} + \epsilon_{s,n} \quad \{\epsilon_{s,n}\} \sim \text{NID}(0, \sigma^2) \quad (4.2.1)$$

where $y_{i,n} = y_{S+i,n-1}$ when $i \leq 0$, and NID denotes normal and independently distributed. The variance and the autocovariance function of this process are periodic. The vector form of PAR is used in many studies, see for example Ghysels and Osborn (2001), Franses (1996), or Troutman (1979). For the case $S = 4$, equation (4.2.1) becomes:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ -\phi_{1,2} & 1 & 0 & 0 \\ 0 & -\phi_{1,3} & 1 & 0 \\ 0 & 0 & -\phi_{1,4} & 1 \end{pmatrix} \begin{pmatrix} y_{1,n} \\ y_{2,n} \\ y_{3,n} \\ y_{4,n} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & \phi_{1,1} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} y_{1,n-1} \\ y_{2,n-1} \\ y_{3,n-1} \\ y_{4,n-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,n} \\ \epsilon_{2,n} \\ \epsilon_{3,n} \\ \epsilon_{4,n} \end{pmatrix}$$

or

$$\begin{aligned} \Phi_0 \mathbf{Y}_n &= \Phi_1 \mathbf{Y}_{n-1} + \mathbf{E}_n \\ \Rightarrow \mathbf{Y}_n &= \Phi_0^{-1} \Phi_1 \mathbf{Y}_{n-1} + \Phi_0^{-1} \mathbf{E}_n \end{aligned} \quad (4.2.2)$$

From equations (4.2.2), we see that a PAR(1) process can be given a VAR(1) representation. In general, a PAR(p) process results in a VAR(P) represen-

tation, where $P = \lceil \frac{p+S-1}{S} \rceil$ and $\lceil \cdot \rceil$ denotes the integer part. For example, for $S = 4$, a PAR(4) will still have a VAR(1) representation with:

$$\Phi_0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -\phi_{1,2} & 1 & 0 & 0 \\ -\phi_{2,3} & -\phi_{1,3} & 1 & 0 \\ -\phi_{3,4} & -\phi_{2,4} & -\phi_{1,4} & 1 \end{pmatrix}$$

$$\Phi_1 = \begin{pmatrix} \phi_{4,1} & \phi_{3,1} & \phi_{1,2} & \phi_{1,1} \\ 0 & \phi_{4,2} & \phi_{3,2} & \phi_{2,2} \\ 0 & 0 & \phi_{4,3} & \phi_{3,3} \\ 0 & 0 & 0 & \phi_{4,4} \end{pmatrix}$$

The general VAR(P) representation is:

$$\begin{aligned} \mathbf{Y}_n &= \Phi_0^{-1} \Phi_1 \mathbf{Y}_{n-1} + \dots + \Phi_0^{-1} \Phi_P \mathbf{Y}_{n-P} + \Phi_0^{-1} \mathbf{E}_n \\ &= \mathbf{A}_1 \mathbf{Y}_{n-1} + \dots + \mathbf{A}_P \mathbf{Y}_{n-P} + \mathbf{U}_n \end{aligned}$$

where $\mathbf{A}_i = \Phi_0^{-1} \Phi_i$, for $(i = 1, \dots, P)$, and $\mathbf{U}_n = \Phi_0^{-1} \mathbf{E}_n$.

A VAR(P) process has a causal stationary representation if the roots of $|\Phi_0 - \Phi_1 z - \dots - \Phi_P z^P|$ are greater than one in absolute value (Hamilton 1994). For a stationary VAR process, the matrix of autocovariances at lag K , $\Gamma(K) = E(\mathbf{Y}_n \mathbf{Y}'_{n-K})$, satisfies the vector Yule-Walker equations

$$\Gamma(K) = \mathbf{A}_1 \Gamma(K-1) + \dots + \mathbf{A}_P \Gamma(K-P) \quad K \geq P$$

For a stationary VAR process, estimation is relatively straightforward (Whittle 1963; Jones and Brelsford 1967) and the standard t and F tests are asymptotically valid.

We define periodic MA models (PMA) in a similar way. A periodic MA(1) process is (Ghysels and Osborn 2001):

$$y_{s,n} = \epsilon_{s,n} + \theta_{1,s} \epsilon_{s-1,n} \quad \{\epsilon_{s,n}\} \sim \text{NID}(0, \sigma^2) \quad (4.2.3)$$

where $\epsilon_{i,n} = \epsilon_{S+i,n-1}$ when $i \leq 0$. As in the PAR case, we can write PMA models in the vector form. For $S = 4$, (4.2.3) becomes:

$$\begin{pmatrix} y_{1,n} \\ y_{2,n} \\ y_{3,n} \\ y_{4,n} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \theta_{1,2} & 1 & 0 & 0 \\ 0 & \theta_{1,3} & 1 & 0 \\ 0 & 0 & \theta_{1,4} & 1 \end{pmatrix} \begin{pmatrix} \epsilon_{1,n} \\ \epsilon_{2,n} \\ \epsilon_{3,n} \\ \epsilon_{4,n} \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & \theta_{1,1} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \epsilon_{1,n-1} \\ \epsilon_{2,n-1} \\ \epsilon_{3,n-1} \\ \epsilon_{4,n-1} \end{pmatrix}$$

or

$$\mathbf{Y}_n = \Theta_1 \mathbf{E}_n + \Theta_2 \mathbf{E}_{n-1} \quad (4.2.4)$$

As in the PAR case, a PMA(q) process will result in a VMA(Q) representation, where $Q = \lceil \frac{q+S-1}{S} \rceil$. The autocovariance function for the model defined by (4.2.4) is:

$$\begin{aligned} \Gamma(0) &= \sigma^2(\Theta_1 \Theta_1' + \Theta_2 \Theta_2') \\ \Gamma(1) &= \sigma^2 \Theta_2 \Theta_1' \\ \Gamma(k) &= \mathbf{0} \quad \text{for } k \geq 2 \end{aligned}$$

As in the univariate case, we can standardise the autocovariance matrix to get the autocorrelation matrix $\mathbf{P}(k) = \mathbf{D}_0^{-1} \Gamma(k) \mathbf{D}_0^{-1}$ where $\mathbf{D}_0^2 = \text{diag}\{\Gamma_{11}(0), \dots, \Gamma_{SS}(0)\}$ and $\Gamma_{ii}(k)$ is the i^{th} diagonal element of $\Gamma(k)$. The autocorrelation matrices for (4.2.3) in the quarterly case are:

$$\begin{aligned} \mathbf{P}(0) &= \begin{pmatrix} 1 & \frac{\theta_{1,2}}{\sqrt{1+\theta_{1,1}^2} \sqrt{1+\theta_{1,2}^2}} & 0 & 0 \\ \frac{\theta_{1,2}}{\sqrt{1+\theta_{1,1}^2} \sqrt{1+\theta_{1,2}^2}} & 1 & \frac{\theta_{1,3}}{\sqrt{1+\theta_{1,2}^2} \sqrt{1+\theta_{1,3}^2}} & 0 \\ 0 & \frac{\theta_{1,3}}{\sqrt{1+\theta_{1,2}^2} \sqrt{1+\theta_{1,3}^2}} & 1 & \frac{\theta_{1,4}}{\sqrt{1+\theta_{1,3}^2} \sqrt{1+\theta_{1,4}^2}} \\ 0 & 0 & \frac{\theta_{1,4}}{\sqrt{1+\theta_{1,3}^2} \sqrt{1+\theta_{1,4}^2}} & 1 \end{pmatrix} \\ \mathbf{P}(1) &= \begin{pmatrix} 0 & 0 & 0 & \frac{\theta_{1,1}}{\sqrt{1+\theta_{1,1}^2} \sqrt{1+\theta_{1,4}^2}} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{aligned} \quad (4.2.5)$$

Shao and Lund (2004) have developed efficient algorithms to compute auto-correlations of PARMA models. The identification of a PAR(p) or a PMA(q) model is not straightforward. In practice, a model selection criterion such as AIC or BIC is used to choose the appropriate order of the model (Franses 1996). Estimation is achieved by computing the exact Gaussian likelihood of PARMA models (Anderson, Meerschaert, and Vecchia 1999; Lund and Basawa 2000; Basawa and Lund 2001)

4.3 Periodic structural times series models

A structural time series model represents the observed series as a sum of un-observed components. Each component is represented explicitly as a stochastic process. We propose a new class of periodic structural time series models. By examining the reduced form we show that PSTSM are observationally equivalent to PIMA models.

4.3.1 Periodic local level model

Consider a time series $\{y_t : t = 1, \dots, SN\}$ with S and N defined in §4.2. A standard decomposition into a trend μ_t and an irregular component ϵ_t is the local level model:

$$\begin{aligned} y_t &= \mu_t + \epsilon_t & \{\epsilon_t\} &\sim \text{NID}(0, \sigma_\epsilon^2) \\ \mu_t &= \mu_{t-1} + \eta_t & \{\eta_t\} &\sim \text{NID}(0, \sigma_\eta^2) \end{aligned} \quad (4.3.1)$$

where $\{\epsilon_t\}$ and $\{\eta_t\}$ are mutually independent. The stationary form of (4.3.1) is:

$$\Delta y_t = \eta_t + \Delta \epsilon_t$$

Allowing the variances in (4.3.1) to be season dependent gives the following model:

$$\begin{aligned} y_{s,n} &= \mu_{s,n} + \epsilon_{s,n} & \{\epsilon_{s,n}\} &\sim \text{NID}(0, \sigma_{\epsilon,s}^2) \\ \mu_{s,n} &= \mu_{s-1,n} + \eta_{s,n} & \{\eta_{s,n}\} &\sim \text{NID}(0, \sigma_{\eta,s}^2) \end{aligned} \quad (4.3.2)$$

where $\mu_{i,n} = \mu_{S+i,n-i}$ for $i \leq 0$. The stationary form of model (4.3.2) is the same as the local level model (4.3.1) but its autocovariance function $c_s(\tau)$ varies with season s .

$$c_s(0) = \sigma_{\eta,s}^2 + \sigma_{\epsilon,s}^2 + \sigma_{\epsilon,s-1}^2 \quad (4.3.3)$$

$$c_s(1) = -\sigma_{\epsilon,s-1}^2 \quad (4.3.4)$$

$$c_s(\tau) = 0 \quad \text{for } \tau \geq 2 \quad (4.3.5)$$

for $s = 1, \dots, S$.

The relationship between PMA and our representation for periodic structural models is established using the vector representation. We rewrite (4.3.2) using the notation of §4.2:

$$\mathbf{Y}_n = \boldsymbol{\mu}_n + \boldsymbol{\epsilon}_n \quad (4.3.6)$$

where $\text{Var}(\boldsymbol{\epsilon}_n) = \boldsymbol{\Sigma}_\epsilon$ is restricted to be diagonal; in the non-periodic case $\boldsymbol{\Sigma}_\epsilon = \sigma_\epsilon^2 \mathbf{I}_S$. The vector representation for the trend is:

$$\begin{pmatrix} 1 & 0 & \dots & \dots & 0 \\ -1 & 1 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \dots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & -1 & 1 \end{pmatrix} \begin{pmatrix} \mu_{1,n} \\ \mu_{2,n} \\ \mu_{3,n} \\ \vdots \\ \mu_{S,n} \end{pmatrix} = \begin{pmatrix} 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} \mu_{1,n-1} \\ \mu_{2,n-1} \\ \vdots \\ \mu_{S,n-1} \end{pmatrix} + \begin{pmatrix} \eta_{1,n} \\ \eta_{2,n} \\ \vdots \\ \eta_{S,n} \end{pmatrix}$$

or

$$\begin{aligned} \mathbf{D}\boldsymbol{\mu}_n &= \boldsymbol{\Phi}\boldsymbol{\mu}_{n-1} + \boldsymbol{\eta}_n \\ \boldsymbol{\mu}_n &= \mathbf{D}^{-1}\boldsymbol{\Phi}\boldsymbol{\mu}_{n-1} + \mathbf{D}^{-1}\boldsymbol{\eta}_n \end{aligned} \quad (4.3.7)$$

Rearranging (4.3.7) yields

$$\begin{pmatrix} \mu_{1,n} \\ \mu_{2,n} \\ \vdots \\ \mu_{S,n} \end{pmatrix} = \begin{pmatrix} 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \begin{pmatrix} \mu_{1,n-1} \\ \mu_{2,n-1} \\ \vdots \\ \mu_{S,n-1} \end{pmatrix} + \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 1 & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} \eta_{1,n} \\ \eta_{2,n} \\ \vdots \\ \eta_{S,n} \end{pmatrix}$$

or

$$\boldsymbol{\mu}_n = \mathbf{T}\boldsymbol{\mu}_{n-1} + \mathbf{R}\boldsymbol{\eta}_n$$

with $\boldsymbol{\eta}_n \sim NID(\mathbf{0}, \boldsymbol{\Sigma}_\eta)$ where $\boldsymbol{\Sigma}_\eta$ is restricted to be diagonal.

\mathbf{T} has one eigenvalue equal to one so $\{\boldsymbol{\mu}_n\}$ is non-stationary (Hamilton 1994).

Applying the difference operator to all elements of $\boldsymbol{\mu}_n$ and noting that \mathbf{T} is

idempotent ($\mathbf{T}^2 = \mathbf{T}$) we have:

$$\begin{aligned} \Delta\boldsymbol{\mu}_n &= \boldsymbol{\mu}_n - \boldsymbol{\mu}_{n-1} \\ &= \mathbf{T}\boldsymbol{\mu}_{n-1} + \mathbf{R}\boldsymbol{\eta}_n - \mathbf{T}\boldsymbol{\mu}_{n-2} - \mathbf{R}\boldsymbol{\eta}_{n-1} \\ &= \mathbf{T}^2\boldsymbol{\mu}_{n-2} + \mathbf{TR}\boldsymbol{\eta}_{n-1} + \mathbf{R}\boldsymbol{\eta}_n - \mathbf{T}\boldsymbol{\mu}_{n-2} - \mathbf{R}\boldsymbol{\eta}_{n-1} \\ &= \mathbf{R}\boldsymbol{\eta}_n + (\mathbf{TR} - \mathbf{R})\boldsymbol{\eta}_{n-1} \\ &= \mathbf{R}\boldsymbol{\eta}_n + \mathbf{W}\boldsymbol{\eta}_{n-1} \end{aligned}$$

where $\mathbf{W} = \mathbf{TR} - \mathbf{R}$. We can then write the vector stationary form of (4.3.6)

as

$$\Delta\mathbf{Y}_n = \mathbf{R}\boldsymbol{\eta}_n + \mathbf{W}\boldsymbol{\eta}_{n-1} + \Delta\boldsymbol{\epsilon}_n$$

Thus $\{\Delta \mathbf{Y}_n\}$ is a vector MA(1) with autocovariance function at lag k , denoted $\Gamma(k)$,

$$\begin{aligned}\Gamma(0) &= \mathbf{R}\Sigma_\eta\mathbf{R}' + \mathbf{W}\Sigma_\eta\mathbf{W}' + 2\Sigma_\epsilon \\ \Gamma(1) &= \mathbf{W}\Sigma_\eta\mathbf{R}' - \Sigma_\epsilon \\ \Gamma(k) &= \mathbf{0} \quad \text{for } k \geq 2\end{aligned}\tag{4.3.8}$$

The non-zero off diagonal elements of the autocorrelation matrices in the quarterly case are:

$$\begin{aligned}\mathbf{P}_{12}(0) &= -\frac{\sigma_{\epsilon_1}^2}{\sqrt{\sigma_{\epsilon_1}^2 + \sigma_{\epsilon_2}^2 + \sigma_{\eta_1}^2}\sqrt{\sigma_{\epsilon_1}^2 + \sigma_{\epsilon_4}^2 + \sigma_{\eta_2}^2}} \\ \mathbf{P}_{23}(0) &= -\frac{\sigma_{\epsilon_2}^2}{\sqrt{\sigma_{\epsilon_1}^2 + \sigma_{\epsilon_4}^2 + \sigma_{\eta_2}^2}\sqrt{\sigma_{\epsilon_2}^2 + \sigma_{\epsilon_3}^2 + \sigma_{\eta_3}^2}} \\ \mathbf{P}_{34}(0) &= -\frac{\sigma_{\epsilon_3}^2}{\sqrt{\sigma_{\epsilon_2}^2 + \sigma_{\epsilon_3}^2 + \sigma_{\eta_3}^2}\sqrt{\sigma_{\epsilon_3}^2 + \sigma_{\epsilon_4}^2 + \sigma_{\eta_4}^2}} \\ \mathbf{P}_{14}(1) &= -\frac{\sigma_{\epsilon_4}^2}{\sqrt{\sigma_{\epsilon_1}^2 + \sigma_{\epsilon_4}^2 + \sigma_{\eta_1}^2}\sqrt{\sigma_{\epsilon_3}^2 + \sigma_{\epsilon_4}^2 + \sigma_{\eta_4}^2}}\end{aligned}\tag{4.3.9}$$

By construction $\mathbf{P}_{11}(0) = \mathbf{P}_{22}(0) = \mathbf{P}_{33}(0) = \mathbf{P}_{44}(0) = 1$. Equating the autocorrelations matrices in equations (4.2.5) and (4.3.9) gives expressions for $\theta_{1,s}$ for $s = 1, \dots, S$. Similar to the non-periodic case (Harvey 1989), the admissible region for $\theta_{1,s}$ in the PSTSM (4.3.2) is $(-1, 0)$ for $s = 1, \dots, S$.

The issue of identifiability is of particular importance in the context of structural models; it is easy to set up models that are not identifiable. We assume normality so the identifiability of the model depends on the form of the autocovariance matrix. Hotta (1989) shows that a sufficient condition for identifiability of an unobserved components ARIMA model is that all the M components have $p_m + d_m \geq q_m + 1$ where p_m , and q_m is the order of

the AR and MA polynomials and d_m is the order of difference for each unobserved component. It is possible to check identifiability for PSTSMs by applying Hotta's result to the reduced form of the components. Alternatively following Harvey (1989, p.207), we can check the identifiability of a periodic structural time series model without invoking the general result. In (4.3.8), we calculated the autocovariance function of the periodic local level model (4.3.2). (4.3.8) provide $2S$ linearly independent equations which give unique solutions for the $2S$ quantities $\sigma_{\epsilon_1}^2, \dots, \sigma_{\epsilon_S}^2$ and $\sigma_{\eta_1}^2, \dots, \sigma_{\eta_S}^2$. This indicates that the model is identifiable. A similar argument establishes identifiability of other periodic structural time series models.

4.3.2 Periodic local linear trend model

We may add a slope component to give the periodic local linear trend model.

$$\begin{aligned} y_{s,n} &= \mu_{s,n} + \epsilon_{s,n} & \{\epsilon_{s,n}\} &\sim \text{NID}(0, \sigma_{\epsilon,s}^2) \\ \mu_{s,n} &= \mu_{s-1,n} + \beta_{s-1,n} + \eta_{s,n} & \{\eta_{s,n}\} &\sim \text{NID}(0, \sigma_{\eta,s}^2) \\ \beta_{s,n} &= \beta_{s-1,n} + \zeta_{s,n} & \{\zeta_{s,n}\} &\sim \text{NID}(0, \sigma_{\zeta,s}^2) \end{aligned} \quad (4.3.10)$$

where $\beta_{i,n} = \beta_{S+i,n-i}$ for $i \leq 0$. The vector representation is then,

$$\begin{aligned} \mathbf{Y}_n &= \boldsymbol{\mu}_n + \boldsymbol{\epsilon}_n \\ \boldsymbol{\mu}_n &= \mathbf{T}\boldsymbol{\mu}_{n-1} + \mathbf{R}\boldsymbol{\beta}_n + \mathbf{R}\boldsymbol{\eta}_n \\ \boldsymbol{\beta}_n &= \mathbf{T}\boldsymbol{\beta}_{n-1} + \mathbf{R}\boldsymbol{\zeta}_n \end{aligned} \quad (4.3.11)$$

with $\boldsymbol{\zeta}_n \sim \text{NID}(\mathbf{0}, \boldsymbol{\Sigma}_\zeta)$. We have:

$$\Delta\boldsymbol{\beta}_n = \mathbf{R}\boldsymbol{\zeta}_n + \mathbf{W}\boldsymbol{\zeta}_{n-1}$$

so that

$$\begin{aligned}
\Delta \boldsymbol{\mu}_n &= \mathbf{T}\boldsymbol{\mu}_{n-1} - \mathbf{T}\boldsymbol{\mu}_{n-2} + \mathbf{R}^2\boldsymbol{\zeta}_n + \mathbf{R}\mathbf{W}\boldsymbol{\zeta}_{n-1} + \mathbf{R}\boldsymbol{\eta}_n - \mathbf{R}\boldsymbol{\eta}_{n-1} \\
&= \mathbf{T}\boldsymbol{\mu}_{n-2} + \mathbf{TR}\boldsymbol{\beta}_{n-1} + \mathbf{TR}\boldsymbol{\eta}_{n-1} - \mathbf{T}\boldsymbol{\mu}_{n-2} + \\
&\quad \mathbf{R}^2\boldsymbol{\zeta}_n + \mathbf{R}\mathbf{W}\boldsymbol{\zeta}_{n-1} + \mathbf{R}\boldsymbol{\eta}_n - \mathbf{R}\boldsymbol{\eta}_{n-1} \\
&= \mathbf{TR}\boldsymbol{\beta}_{n-1} + \mathbf{R}\boldsymbol{\eta}_n + \mathbf{W}\boldsymbol{\eta}_{n-1} + \mathbf{R}^2\boldsymbol{\zeta}_n + \mathbf{R}\mathbf{W}\boldsymbol{\zeta}_{n-1}
\end{aligned}$$

It follows that:

$$\begin{aligned}
\Delta^2 \boldsymbol{\mu}_n &= \mathbf{TR}^2\boldsymbol{\zeta}_{n-1} + \mathbf{TR}\mathbf{W}\boldsymbol{\zeta}_{n-2} + \mathbf{R}\boldsymbol{\eta}_n - \mathbf{R}\boldsymbol{\eta}_{n-1} + \mathbf{W}\boldsymbol{\eta}_{n-1} - \\
&\quad \mathbf{W}\boldsymbol{\eta}_{n-2} + \mathbf{R}^2\boldsymbol{\zeta}_n - \mathbf{R}^2\boldsymbol{\zeta}_{n-1} + \mathbf{R}\mathbf{W}\boldsymbol{\zeta}_{n-1} - \mathbf{R}\mathbf{W}\boldsymbol{\zeta}_{n-2} \\
&= \mathbf{R}\boldsymbol{\eta}_n + (\mathbf{W} - \mathbf{R})\boldsymbol{\eta}_{n-1} - \mathbf{W}\boldsymbol{\eta}_{n-2} + \\
&\quad \mathbf{R}^2\boldsymbol{\zeta}_n + (\mathbf{TR}^2 - \mathbf{R}^2 + \mathbf{R}\mathbf{W})\boldsymbol{\zeta}_{n-1} + (\mathbf{TR}\mathbf{W} - \mathbf{R}\mathbf{W})\boldsymbol{\zeta}_{n-2} \\
&= \mathbf{R}\boldsymbol{\eta}_n + (\mathbf{W} - \mathbf{R})\boldsymbol{\eta}_{n-1} - \mathbf{W}\boldsymbol{\eta}_{n-2} + \\
&\quad \mathbf{R}^2\boldsymbol{\zeta}_n + (\mathbf{W}\mathbf{R} + \mathbf{R}\mathbf{W})\boldsymbol{\zeta}_{n-1} + \mathbf{W}^2\boldsymbol{\zeta}_{n-2}
\end{aligned}$$

The stationary form is then:

$$\Delta^2 \mathbf{Y}_n = \Delta^2 \boldsymbol{\mu}_n + \boldsymbol{\epsilon}_n - 2\boldsymbol{\epsilon}_{n-1} + \boldsymbol{\epsilon}_{n-2}$$

$\Delta^2 \mathbf{y}_n$ is a vector MA(2) with autocovariance function at lag k , denoted as $\Gamma(k)$:

$$\begin{aligned}
\Gamma(0) &= 2\mathbf{R}\Sigma_\eta\mathbf{R}' + 2\mathbf{W}\Sigma_\eta\mathbf{W}' - \mathbf{W}\Sigma_\eta\mathbf{R}' - \mathbf{R}\Sigma_\eta\mathbf{W}' + \\
&\quad \mathbf{R}^2\Sigma_\zeta\mathbf{R}'^2 + \mathbf{W}\mathbf{R}\Sigma_\zeta\mathbf{R}'\mathbf{W}' + \mathbf{W}\mathbf{R}\Sigma_\zeta\mathbf{W}'\mathbf{R}' + \\
&\quad \mathbf{R}\mathbf{W}\Sigma_\zeta\mathbf{R}'\mathbf{W}' + \mathbf{R}\mathbf{W}\Sigma_\zeta\mathbf{W}'\mathbf{R}' + \mathbf{W}^2\Sigma_\zeta\mathbf{W}'^2 + 6\Sigma_\epsilon \\
\Gamma(1) &= 2\mathbf{W}\Sigma_\eta\mathbf{R}' - \mathbf{R}\Sigma_\eta\mathbf{R}' - \mathbf{W}\Sigma_\eta\mathbf{W}' + \\
&\quad \mathbf{W}\mathbf{R}\Sigma_\zeta\mathbf{R}'^2 + \mathbf{R}\mathbf{W}\Sigma_\zeta\mathbf{R}'^2 + \mathbf{W}^2\Sigma_\zeta\mathbf{R}'\mathbf{W}' + \mathbf{W}^2\Sigma_\zeta\mathbf{W}'\mathbf{R}' - 4\Sigma_\epsilon \\
\Gamma(2) &= -\mathbf{W}\Sigma_\eta\mathbf{R}' + \mathbf{W}^2\Sigma_\zeta\mathbf{R}'^2 + \Sigma_\epsilon \\
\Gamma(k) &= \mathbf{0} \quad \text{for } k \geq 3 \quad .
\end{aligned}$$

$\Gamma(2)$ is an upper triangular matrix which corresponds to $\{\Delta_S^2 y_t\}$ being a PMA(2S).

4.3.3 Periodic basic structural model

Adding a seasonal component γ_n to the periodic local level model (4.3.2), or in the periodic local linear trend model (4.3.10) results in a periodic version of the basic structural model (BSM), (Harvey 1989). The vector representation is:

$$\mathbf{Y}_n = \boldsymbol{\mu}_n + \boldsymbol{\gamma}_n + \boldsymbol{\epsilon}_n \quad (4.3.12)$$

We examine two possible representations of the seasonal component below.

Harrison-Stevens seasonality model

For the seasonal component, it is natural to use the Harrison and Stevens (1976) seasonal model. We re-write (2.3.16) by stacking all the seasons in a vector form:

$$\gamma_n = \gamma_{n-1} + \omega_n \quad (4.3.13)$$

where ω_n is a zero-mean process with variance (?):

$$\text{Var}(\omega_n) = \Omega = \mathbf{D} - \frac{1}{\mathbf{i}'_s \mathbf{D} \mathbf{i}_s} \mathbf{D} \mathbf{i}_s \mathbf{i}'_s \mathbf{D} \quad (4.3.14)$$

where $\mathbf{D} = \text{diag}\{\sigma_{\omega,1}^2, \dots, \sigma_{\omega,S}^2\}$ and $\mathbf{i}_s = [1, 1, \dots, 1]'$ is an $S \times 1$. The variance-covariance matrix Ω enforces the constraint that $\mathbf{i}'_s \text{Var}(\omega_t) = 0$. As we saw in §2.3.3 the model enforces the constraint that seasonality adds up to zero within a year. $\{\Delta \gamma_n\}$ is a vector MA(1) process with autocovariance:

$$\begin{aligned} \Gamma(0) &= \Omega \\ \Gamma(1) &= \mathbf{V} \\ \Gamma(k) &= \mathbf{0} \quad \text{for } k \geq 2 \end{aligned}$$

The elements of \mathbf{V} are defined as follows:

$$\mathbf{V}_{i,j} = \begin{cases} \Omega_{i,j} & \text{if } i < j \\ 0 & \text{if } i \geq j \end{cases}$$

A HS seasonal component in the BSM (4.3.12), results in $\{\Delta^2 \mathbf{Y}_n\}$ being a stationary vector MA(2) as is the case of model (4.3.10). Similarly, adding a HS seasonal component in the periodic local level model (4.3.2) still results in a vector MA(1) process. However there are fewer non-zero elements in

the autocorrelation matrices than in the non-seasonal case. In fact the autocorrelation matrices of the local level model with HS seasonality correspond to $\Delta_s y_t$ being a PMA(S). Equating the two autocorrelation matrices we get exact expressions for the $S \times S$ unknown $\theta_{i,s}$.

Dummy seasonality

An alternative model for γ_n is the dummy seasonality model we introduced in §2.3.1. In this case, stacking up the equations (2.3.2) for all the seasons gives:

$$\begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \ddots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 1 & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} \gamma_{1,n} \\ \gamma_{2,n} \\ \vdots \\ \gamma_{S,n} \end{pmatrix} = - \begin{pmatrix} 0 & 1 & \dots & 1 \\ 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & 1 \\ 0 & 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} \gamma_{1,n-1} \\ \gamma_{2,n-1} \\ \vdots \\ \gamma_{S,n-1} \end{pmatrix} + \begin{pmatrix} \omega_{1,n} \\ \omega_{2,n} \\ \vdots \\ \omega_{S,n} \end{pmatrix}$$

or in vector notation:

$$\begin{aligned} \mathbf{R}\gamma_n &= -\mathbf{W}\gamma_{n-1} + \boldsymbol{\omega}_n \\ \gamma_n &= -\mathbf{R}^{-1}\mathbf{W}\gamma_{n-1} + \mathbf{R}^{-1}\boldsymbol{\omega}_n \end{aligned} \quad (4.3.15)$$

with $\boldsymbol{\omega}_n \sim NID(\mathbf{0}, \boldsymbol{\Sigma}_\omega)$ where $\boldsymbol{\Sigma}_\omega$ is restricted to be diagonal. In the time invariant case $\boldsymbol{\Sigma}_\omega = \sigma_\omega^2 \mathbf{I}$. (4.3.15) becomes:

$$\begin{pmatrix} \gamma_{1,n} \\ \gamma_{2,n} \\ \gamma_{3,n} \\ \vdots \\ \gamma_{S,n} \end{pmatrix} = - \begin{pmatrix} 0 & 1 & 1 & \dots & 1 \\ 0 & -1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \dots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & -1 \end{pmatrix} \begin{pmatrix} \gamma_{1,n-1} \\ \gamma_{2,n-1} \\ \gamma_{3,n-1} \\ \vdots \\ \gamma_{S,n-1} \end{pmatrix} + \begin{pmatrix} 1 & 0 & \dots & \dots & 0 \\ -1 & 1 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \dots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & -1 & 1 \end{pmatrix} \begin{pmatrix} \omega_{1,n} \\ \omega_{2,n} \\ \omega_{3,n} \\ \vdots \\ \omega_{4,n} \end{pmatrix}$$

or

$$\gamma_n = -\mathbf{J}\gamma_{n-1} + \mathbf{D}\boldsymbol{\omega}_n$$

Since \mathbf{J} has $S - 1$ eigenvalues equal to -1 , γ_n is not stationary. Applying the difference operator to all elements of γ_n and noting that $\mathbf{J}^2 = -\mathbf{J}$ we have:

$$\begin{aligned}
\Delta\gamma_n &= \gamma_n - \gamma_{n-1} \\
&= -\mathbf{J}\gamma_{n-1} + \mathbf{D}\omega_n + \mathbf{J}\gamma_{n-2} - \mathbf{D}\omega_{n-1} \\
&= \mathbf{J}^2\gamma_{n-2} - \mathbf{J}\mathbf{D}\omega_{n-1} + \mathbf{D}\omega_n + \mathbf{J}\gamma_{n-2} - \mathbf{D}\omega_{n-1} \\
&= \mathbf{D}\omega_n - (\mathbf{J} + \mathbf{I}_s)\mathbf{D}\omega_{n-1} \\
&= \mathbf{D}\omega_n - \Phi\omega_{n-1}
\end{aligned}$$

since $(\mathbf{J} + \mathbf{I}_s)\mathbf{D} = \Phi$ where Φ is defined in (4.3.7). Thus, $\{\Delta\gamma_n\}$ is a vector MA(1) process with autocovariance function

$$\begin{aligned}
\Gamma(0) &= \mathbf{D}\Sigma_\omega\mathbf{D}' + \Phi\Sigma_\omega\Phi' \\
\Gamma(1) &= -\Phi\Sigma_\omega\mathbf{D}' \\
\Gamma(k) &= \mathbf{0} \quad \text{for } k \geq 2
\end{aligned}$$

As in the case of Harrison-Stevens, adding a dummy seasonal component in the local level model (4.3.1), still gives a vector MA(1) process.

4.4 Applications

We propose a strategy to determine the appropriate periodic structural time series model. Franses (1996) suggests the use of model selection criteria such as AIC or BIC for determining the order of a PAR model. We adopt a similar approach for PSTSM. As a first step we select the appropriate time-invariant STSM by using diagnostic tests suggested by Harvey (1989). We

use the standardized one-step ahead prediction errors $e_t = \frac{v_t}{\sqrt{F_t}}$, where v_t is the one-step ahead prediction error and F_t is its variance. If the model is correctly specified and all the parameters known, then $\{e_t\} \sim \text{NID}(0, 1)$. We use common tests for normality and serial correlation such as those suggested by Bowman and Shenton (1975) and Ljung and Box (1978). The exact likelihood for periodic structural time series is calculated using the multivariate Kalman filter (Harvey 1989). The efficiency of the Kalman filter makes likelihood ratio tests a convenient tool for inference. The algorithm used for estimation and signal extraction is implemented in Ox (Doornik 1998) using SsfPack (Koopman, Shephard, and Doornik 1998). We test for seasonal heteroscedasticity in every season separately using a likelihood ratio test and then use the AIC to choose the appropriate combination of periodic variances to formulate a PSTSM. Finally the chosen PSTSM model is checked for normality and independence of the residuals using the tests discussed in §2.8.

We compare the performance of PAR models with the PSTSM on a data set of quarterly macroeconomic variables used by Franses (1996). They include several macroeconomic indicators from UK, USA, Canada, and Germany, which are plotted in figures (4.1)-(4.1). We use a logarithmic transformation to all the series except for the Canadian unemployment. In that monograph, the series are scrutinized for periodicity in the AR parameters using a battery of tests. The author arrived at an optimal order of periodic AR using the Schwarz information criterion and an F -test for the significance of $\phi_{p+1,s}$ parameters. Alternatively, the Akaike information criterion can be used.

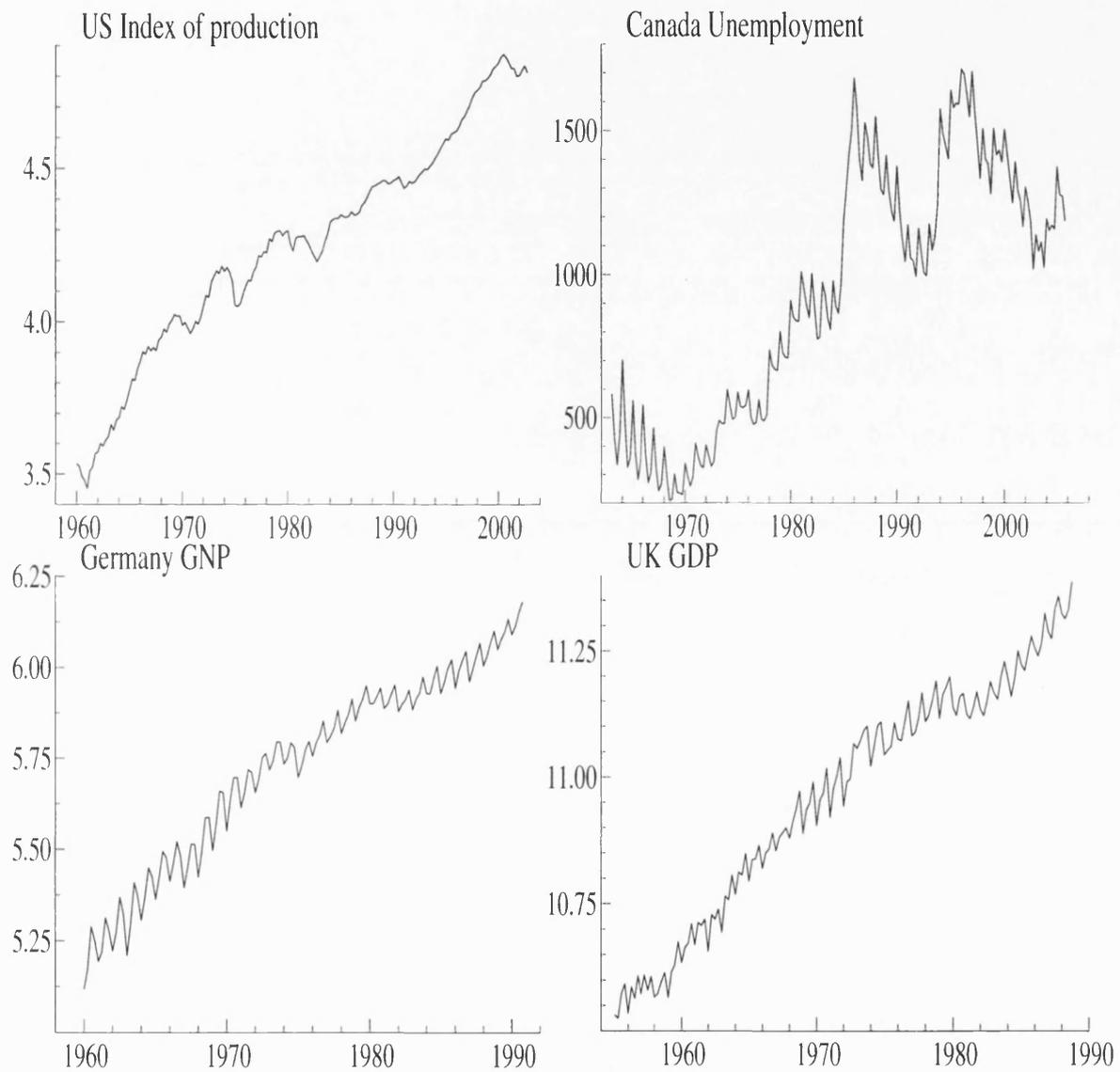


Figure 4.1: Time series used in the application of periodic models (1)

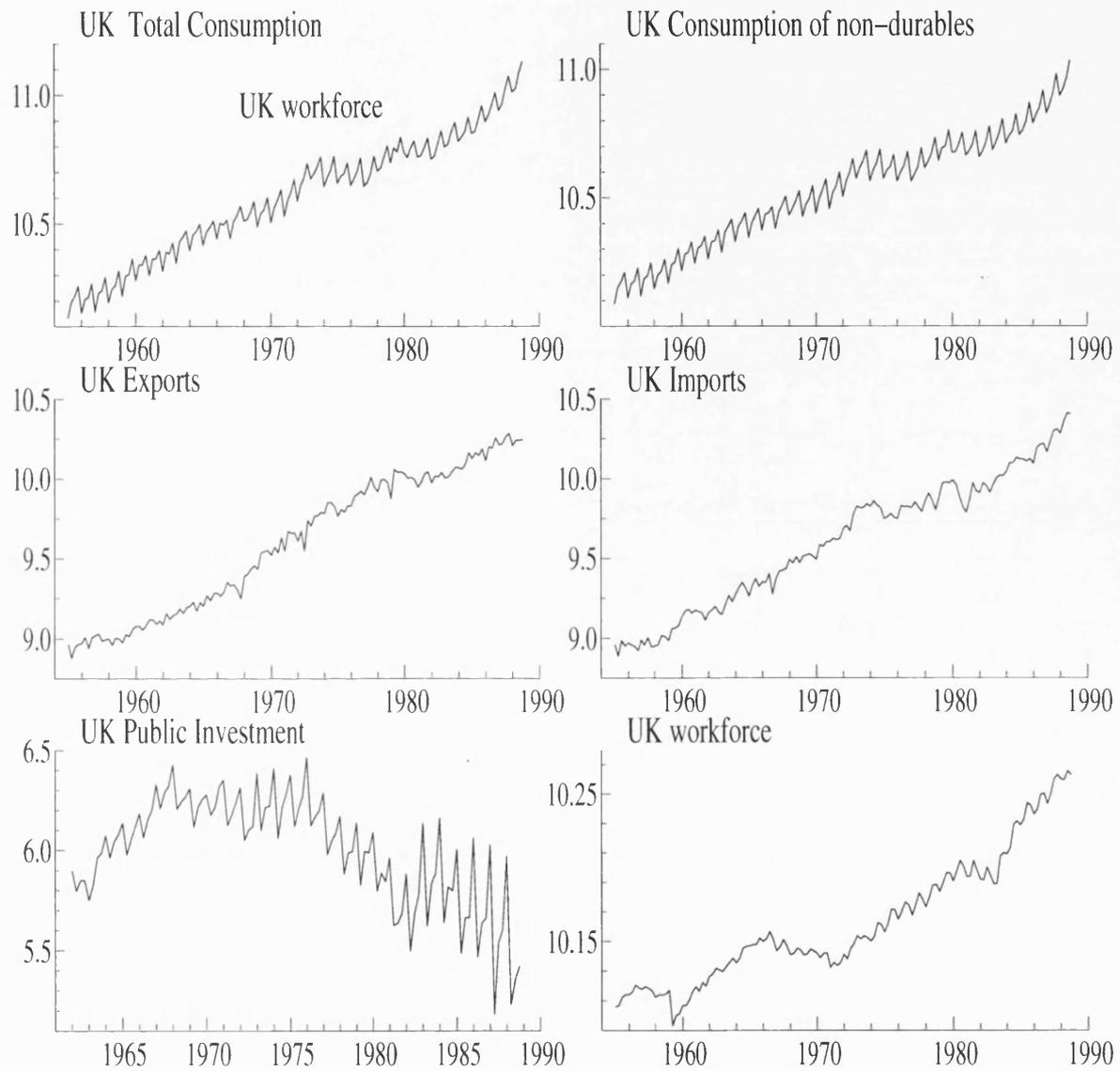


Figure 4.2: Time series used in the application of periodic models (2)

We use the same p as Franses (1996) who also considered PAR models that include deterministic seasonal trends. PAR models for all series were cast in state-space form and estimated using Ox (Doornik 1998) by maximum likelihood using the Kalman filter. We give results for the chosen PAR models with and without trend. Novales and de Frutto (1997) report that the forecasting performance of PAR models improves considerably by imposing non-periodic coefficients across some seasons so we included the results of a constrained version not considered by Franses (1996). Column C in Table 4.1 shows the seasons that have varying coefficients in the constrained PAR model. We used the AIC to select the constrained model. We fitted all different combinations of constraints and chose the model that best fitted the data according to AIC. For all the series except the unemployment series for Canada, the constrained PAR has lower AIC than the unconstrained version. For the PSTSM, we first selected between a local level model (4.3.2) and a local linear trend model (4.3.10) using the AIC. We then select which quarters have different variance by fitting all the different combinations of seasonal heteroscedastic models and select the model that minimises the AIC. As explained in §2.8, the AIC is comparable across the different models and gives a fair comparison between PAR and PSTSM models. For all the series, except the Candian unemployment series, PSTSM fitted the data better than the best PAR, in terms of AIC

The best model from Table 4.1 is fitted without the final year of data and forecasts generated for the deleted observations. This experiment is repeated removing the last two and then the last three years of data. Table 4.2 shows

Table 4.1: Comparison of PAR and PSTSM (log is applied to all the series except CAN Unemployment)

Variable	p ^a	C ^b	S ^c	AIC _{PAR} ^d	AIC _{PAR} ^e	AIC _{PAR} ^f	AIC _{PSTSM}
USA Ind. production	2	2,4	4	-6.73	-6.71	-6.73 ^d	-9.89 ^h
CAN Unemployment ⁱ	4	3	^j	9.57	9.65	9.64 ^d	17.33 ^g
DEU GNP,	2	2,4	2	-6.69	-6.68	-6.73 ^e	-12.23 ^g
UK GDP	2	2,3	3	-6.35	-6.34	-6.47 ^d	-8.86 ^h
UK consumption	1	2	3	-6.82	-6.80	-6.95 ^d	-9.36 ^h
UK cons. nondur.	1	2	3,4	-7.47	-7.43	-7.61 ^d	-10.47 ^g
UK Exports	2	2	3	-4.88	-4.86	-4.99 ^e	-7.36 ^h
UK Imports	1	2,4	3	-4.95	-4.98	-5.08 ^e	-9.30 ^g
UK pub. investment	2	1	2	-2.53	-2.62	-2.73 ^e	-4.94 ^h
UK workforce	2	1	1,2,3,4	-9.43	-9.4	-9.61 ^d	-9.62 ^h

^a Order of PAR model

^b Seasons with varying coefficients in the PAR model

^c Seasons with different variance in the PSTSM model

^d Model without trend

^e Model with Trend

^f Constrained PAR

^g Fixed Level+Stochastic Slope

^h Stochastic Level+ Fixed Slope

ⁱ log-trasformation is not applied in this series

^j No heteroscedasticity

Table 4.2: Out of Sample Comparison of PAR and STSM (RMSE)

Variable	1 Year		2 Years		3 Years	
	PAR	PSTSM	PAR	PSTSM	PAR	PSTSM
USA, Industrial production	0.31	0.06	0.35	0.09	0.02	0.09
CAN Unemployment ^a	4.34	61.18	3.5	129.39	59.1	344.3
DEU, Real GNP	0.02	0.02	0.02	0.02	0.02	0.04
UK GDP	0.16	0.02	0.22	0.01	0.03	0.01
UK total consumption	0.099	0.008	0.02	0.01	0.01	0.02
UK consumer nondurables	0.10	0.009	0.10	0.01	0.01	0.02
UK Exports	0.03	0.05	0.03	0.08	0.03	0.03
UK Imports	0.04	0.02	0.05	0.03	0.05	0.19
UK public investment	0.19	0.10	0.24	0.16	0.24	0.18
UK workforce	0.006	0.008	0.005	0.004	0.005	0.006

^a log-trasformation is not applied in this series

the root mean square error for the out-of-sample forecasts. The results suggest that, at least for the one and two year ahead forecasts, PSTSM outperforms PAR. This suggests that, in many instances, PIMA may provide a better representation than PAR for the correlation structure of economic series. For the longest forecast horizon, PAR models perform marginally better in terms of out of sample RMSE. Some of the forecasts may be improved by including other cyclical components or exogenous variables, but for the purposes of this thesis we concentrated in showing a fair comparison between PSTSM and PAR models. We conclude that, for these series, PSTSM produce considerable gains in accuracy over PAR models, at least for short and medium term forecasts.

4.5 Conclusions

Periodic processes where the coefficients change with the season represent a real feature of economic time series. The most widely used framework is the periodic autoregression (PAR). We advocate representation of periodic processes in the structural time series framework. We established that periodic structural models have a vector integrated moving average representation. Although vector moving average and vector integrated moving average models have been investigated as part of VARIMA class of models, they are little used in practice. We have shown that a class of models with VIMA correlation structure provide parsimonious models for univariate series with periodic second moments. Moreover, in practical applications, the structural framework provides greater insight into the nature of the series than PAR models. PSTSM relate the seasonality in the autocovariance function to specific unobserved components. In comparison, PAR models parameters are not readily interpretable. We compare the forecasting accuracy of PSTSM with PAR models. PSTSM produced better forecasts both within and out of sample for the majority of the series concerned. We conclude that structural time series models are a natural framework for modelling periodic processes both in terms of interpretability and forecasting accuracy.

Chapter 5

Contemporaneous aggregation of time series

5.1 Introduction

Many economic time series can be broken down into components, such as regions or different types of economic activity. For example, European Monetary Union (EMU) GDP can be disaggregated into the GDPs of the member states. If the individual components are observed, then the aggregate series can be constructed by adding up these components, as is the case in the EMU GDP. If the aim is to forecast the aggregate series then the question arises about forecasting individual series and then aggregating or directly forecasting the aggregate series. This can be summarised as whether to aggregate the forecast or forecast the aggregate. In fact, the choice of the aggregate series is arbitrary. Each series in a hierarchical dataset with n time series can be seen as linear combination of the $n - 1$ other time series. Hotta and

Vasconcellos (1999) consider the problem of temporal aggregation, such as turning a monthly series into a quarterly. We concentrate on the problem of sectional aggregation; adding up different series observed in the same time period. There is an extensive literature on this topic, see Lütkepohl (1987) and references therein, but very few practical solutions.

We devise a methodology to take an appropriate decision for each series. Using the same methodology, we can also decide what is the best way of aggregating a dataset in order to get minimum MSE forecasts. There are three forecasting methods that we can distinguish:

1. model each individual series separately and then aggregate the forecasts
2. model all the series with a multivariate time series model and then aggregate the forecasts; we use a seemingly unrelated time series equation model (SUTSE), see Harvey (1989)
3. model the aggregate series through a univariate model.

Lütkepohl (1987) shows, using ARMA models, that 2 is at least as good as 1 and 3 in terms of MSE. He also shows that the gains between the predictors vanishes for long-range forecasting since all three prediction MSE matrices converge to the MSE of method 3. Wei and Abraham (1981) show that the comparison of 1 and 3 in terms of MSE depends on the structure of the component series and the method of aggregation. Giacomini and Granger (2004) add a fourth method by adding a spatial element and show that it is at least as good as 2. In this chapter, we consider only models with a

time element and we do not borrow information from other related series or dimensions, such as space. Related work and special cases results can be found in Granger and Morris (1976), Rose (1977) and Tiao and Guttman (1980).

In §5.2 we define the problem of forecasting aggregated time series. We look at the conditions under which a structural time series model is identifiable. We show that simple aggregation of identifiable models does not ensure that the aggregate model will be identifiable. In the following section, we compare forecasts under the assumption of no estimation error. We show that we can use quantities derived by simple manipulations of the Kalman filter to compare the MSEs of the three methodologies. In §5.4 we compare the forecasts of the aggregated series when the parameters of the model are unknown and need to be estimated. We provide the recursions that are required to derive the asymptotic conditional variance of the one-step ahead prediction error. In §5.5, we apply our methodology in the UK motor vehicle production dataset. This small dataset includes three series, and we investigate different ways of aggregation in order to get the best forecasts for all series.

5.2 Aggregation of time series

Let p be the number of time series and let $\mathbf{y}_t = (y_{1,t}, \dots, y_{p,t})'$ denote the observations of p time series at time t . Then the aggregate at time t is $y_t^A = \mathbf{w}\mathbf{y}_t$ where $\mathbf{w} = (w_1 \dots w_p)$ is a vector of weights which we assume to

be constant over time. We also use $\mathbf{Y}_{j,t} = (y_{j,1}, \dots, y_{j,t})'$ to denote the vector of past observations of the j^{th} series, and $\mathbf{Y}_t = (\mathbf{Y}_{1,t}, \dots, \mathbf{Y}_{p,t})$ the vector of past observations of all p time series. We make the following assumption about the formulation of the model.

Assumption 5.2.1. *All time series $y_{j,t}$ can be written in a state-space form*

We also need to ensure that the aggregate series has an identifiable state-space representation.

Lemma 5.2.2. *A sufficient condition for an aggregate time series y_t^A to have a state space representation is that all p time series $y_{j,t}$ have a state-space representation.*

Proof. A time series has a state-space representation if there exists a state-space models for $y_{j,t}$ specified by the following equation

$$\begin{aligned} y_{j,t} &= \mathbf{Z}_{j,t} \boldsymbol{\alpha}_{j,t} + \epsilon_{j,t} & \epsilon_{j,t} &\sim N(0, H_{j,t}) \\ \boldsymbol{\alpha}_{j,t+1} &= \mathbf{T}_{j,t} \boldsymbol{\alpha}_{j,t} + \mathbf{R}_{j,t} \boldsymbol{\eta}_{j,t} & \boldsymbol{\eta}_{j,t} &\sim N(\mathbf{0}, \mathbf{Q}_{j,t}) \end{aligned}$$

for $j = 1, \dots, p$. The dimension of each state $\boldsymbol{\alpha}_{j,t}$ is m_j . Then the aggregate

series has the following form:

$$\begin{aligned}
 y_t^A = \mathbf{w}y_t &= \begin{pmatrix} w_1 & \dots & w_p \end{pmatrix} \begin{pmatrix} \mathbf{Z}_{1,t} & 0 & \dots & 0 \\ 0 & \mathbf{Z}_{2,t} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \mathbf{Z}_{p,t} \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha}_{1,t} \\ \vdots \\ \boldsymbol{\alpha}_{p,t} \end{pmatrix} \\
 &+ \begin{pmatrix} w_1 & \dots & w_p \end{pmatrix} \begin{pmatrix} \epsilon_{1,t} \\ \vdots \\ \epsilon_{p,t} \end{pmatrix}
 \end{aligned}$$

which shows that y_t^A has a state-space representation. If $\epsilon_{j,t}$ are uncorrelated in time so that $E(\epsilon_{j,t}\epsilon_{l,t-k}) = 0$ for $k = 1, \dots, t-1$ then:

$$\begin{aligned}
 y_t^A &= \mathbf{Z}_t^A \boldsymbol{\alpha}_t + \epsilon_t^A & \epsilon_t^A &\sim N(0, H_t^A) \\
 \boldsymbol{\alpha}_{t+1} &= \mathbf{T}_t^A \boldsymbol{\alpha}_t + \mathbf{R}_t^A \boldsymbol{\eta}_t^A & \boldsymbol{\eta}_t^A &\sim N(\mathbf{0}, \mathbf{Q}_t^A)
 \end{aligned} \tag{5.2.1}$$

where

$$\mathbf{Z}_t^A = \begin{pmatrix} w_1 \mathbf{Z}_{1,t} & w_2 \mathbf{Z}_{2,t} & \dots & w_p \mathbf{Z}_{p,t} \end{pmatrix}$$

$$\boldsymbol{\alpha}'_t = \begin{pmatrix} \boldsymbol{\alpha}_{1,t} & \boldsymbol{\alpha}_{2,t} & \dots & \boldsymbol{\alpha}_{p,t} \end{pmatrix}$$

$$\mathbf{T}_t^A = \begin{pmatrix} \mathbf{T}_{1,t} & 0 & \dots & 0 \\ 0 & \mathbf{T}_{2,t} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \mathbf{T}_{p,t} \end{pmatrix}$$

and $H_t^A = \sum_{j=1}^p w_j^2 H_{j,t}$. The dimension(dim) of \mathbf{Z}_t^A and $\boldsymbol{\alpha}'_t$ is $1 \sum_{j=1}^p m_j$, and of $\dim \mathbf{T}_t^A = \sum_{j=1}^p m_j \times \sum_{j=1}^p m_j$. □

We define the three forecasting methods that we described in the introduction in terms of conditioned estimates of the state.

$$(\mathbf{E}(\boldsymbol{\alpha}_{1,t} | \mathbf{Y}_{1,t})' \dots \mathbf{E}(\boldsymbol{\alpha}_{p,t} | \mathbf{Y}_{p,t})') = \mathbf{a}_t^{(1)} \quad \text{Method 1(univariate)}$$

$$(\mathbf{E}(\boldsymbol{\alpha}_{1,t} | \mathbf{Y}_t)' \dots \mathbf{E}(\boldsymbol{\alpha}_{p,t} | \mathbf{Y}_t)') = \mathbf{a}_t^{(2)} \quad \text{Method 2(multivariate)}$$

$$\mathbf{E}(\boldsymbol{\alpha}_t^A | \mathbf{w} \mathbf{Y}_t) = \mathbf{a}_t^{(3)} \quad \text{Method 3(Direct)}$$

For methods 1 and 2, the state-space formulation takes the form of 5.2.1, so that $\dim(\mathbf{a}_t^{(1)}) = \dim(\mathbf{a}_t^{(2)}) = \dim(\mathbf{a}_t)$. The difference between methods 1 and 2 for the structural time series models are in the matrices \mathbf{H}_t^A and \mathbf{Q}_t^A ,

which for the univariate case are block diagonal, while for the multivariate case they have non-zero elements in the off diagonal. This has an effect in the filter derived matrices \mathbf{P}_t and \mathbf{F}_t which are block diagonal in the univariate case, while they are not in the multivariate case.

To illustrate the point, we have two series of which one follows the local level model and the other the local linear trend model. For simplicity we use $w_1 = w_2 = 1$:

$$\begin{aligned} y_{1,t} &= \mu_{1,t} + \epsilon_{1,t} \\ y_{2,t} &= \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} \mu_{2,t} \\ \beta_{2,t} \end{pmatrix} + \epsilon_{2,t} \end{aligned}$$

$$\begin{aligned} \mu_{1,t} &= \mu_{1,t-1} + \eta_{1,t} \\ \begin{pmatrix} \mu_{2,t} \\ \beta_{2,t} \end{pmatrix} &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{2,t-1} \\ \beta_{2,t-1} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \eta_{2,t} \\ \zeta_{2,t} \end{pmatrix} \end{aligned}$$

When we aggregate the two series $y_{1,t}$ and $y_{2,t}$ we get y_t^A

$$\begin{aligned} y_t^A &= \mu_t^A + \epsilon_t^A \\ \epsilon_t^A &= \epsilon_{1,t} + \epsilon_{2,t} \\ \mu_t^A &= \mu_{1,t-1} + \mu_{2,t-1} + \beta_{2,t} + \eta_{1,t} + \eta_{2,t} \\ \beta_{2,t} &= \beta_{2,t-1} + \zeta_{2,t} \end{aligned}$$

where $\epsilon_{j,t} \sim \text{NID}(0, \sigma_{\epsilon_j}^2)$ for $j = 1, 2$, $\eta_{j,t} \sim \text{NID}(0, \sigma_{\eta_j}^2)$ for $j = 1, 2$, and $\zeta_{2,t} \sim \text{NID}(0, \sigma_{\zeta_2}^2)$ are mutually independent. \mathbf{Z}_t^A and \mathbf{T}_t^A are the same for both methods 1 and 2.

$$\mathbf{Z}_t^A = \begin{pmatrix} 1 & 1 & 0 \end{pmatrix} \quad \mathbf{T}_t^A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

The matrices of the hyperparameters \mathbf{H}_t^A and \mathbf{Q}_t^A are different for each method

Method 1

$$\mathbf{H}_t^A = \begin{pmatrix} \sigma_{\epsilon_1}^2 & 0 \\ 0 & \sigma_{\epsilon_2}^2 \end{pmatrix} \quad \mathbf{Q}_t^A = \begin{pmatrix} \sigma_{\eta_1}^2 & 0 & 0 \\ 0 & \sigma_{\eta_2}^2 & 0 \\ 0 & 0 & \sigma_{\zeta_2}^2 \end{pmatrix}$$

Method 2

$$\mathbf{H}_t^A = \begin{pmatrix} \sigma_{\epsilon_1}^2 & \sigma_{\epsilon_{12}} \\ \sigma_{\epsilon_{12}} & \sigma_{\epsilon_2}^2 \end{pmatrix} \quad \mathbf{Q}_t^A = \begin{pmatrix} \sigma_{\eta_1}^2 & \sigma_{\eta_{12}} & \sigma_{\eta_1\zeta_2} \\ \sigma_{\eta_{12}} & \sigma_{\eta_2}^2 & 0 \\ \sigma_{\eta_1\zeta_2} & 0 & \sigma_{\zeta_2}^2 \end{pmatrix}$$

with $\mathbf{R}_t^A = \mathbf{I}_3$ for both methods. As noted before, *Method 2* is equivalent to a seemingly unrelated time series equation (SUTSE) model (Harvey 1989).

The state-space form method 3 is:

$$\begin{aligned} y_t^A &= \mathbf{Z}_t^\dagger \boldsymbol{\alpha}_t^{(3)} + \epsilon_t^\dagger & \epsilon_t^\dagger &\sim N(0, H_t^\dagger) \\ \boldsymbol{\alpha}_{t+1}^{(3)} &= \mathbf{T}_t^\dagger \boldsymbol{\alpha}_t^{(3)} + \mathbf{R}_t^\dagger \boldsymbol{\eta}_t^\dagger & \boldsymbol{\eta}_t^\dagger &\sim N(\mathbf{0}, \mathbf{Q}_t^\dagger) \end{aligned}$$

The form of \mathbf{Z}_t^\dagger and consequently the dimension of the state vector $\boldsymbol{\alpha}_t^{(3)}$ depends on the form of the model of the original series. In the structural time series framework, the aggregate model would be the one that encompasses all models of the individual series. This means that \mathbf{Z}_t^\dagger certainly is not equal to \mathbf{Z}_t^A , and does not have to be equal to $\mathbf{Z}_{j,t}$. In the case of two series where one follows the local level model, and the other has only a seasonal component, we model the aggregate series using the Basic Structural Model, so $\mathbf{Z}_t^\dagger \neq \mathbf{Z}_{j,t}$ for $j = 1, 2$. On the other hand, for our example, with a local level plus a local linear trend model, the aggregate series is described by a local linear trend model so that $\mathbf{Z}_t^\dagger = \mathbf{Z}_{2,t}$ and $\mathbf{T}_t^\dagger = \mathbf{T}_{2,t}$ so that

$$\mathbf{Z}_t^\dagger = \begin{pmatrix} 1 & 0 \end{pmatrix} \quad \mathbf{T}_t^\dagger = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

and the matrices of the hyperparameters for method 3 are:

$$\begin{aligned} H_t^\dagger &= \sigma_{\epsilon^\dagger}^2 \\ \mathbf{Q}_t^\dagger &= \begin{pmatrix} \sigma_{\eta^\dagger}^2 & 0 \\ 0 & \sigma_{\zeta^\dagger}^2 \end{pmatrix} \\ \mathbf{R}_t^\dagger &= \mathbf{I}_2 \end{aligned}$$

We note that $H_t^\dagger \neq H_{2,t}$ and $\mathbf{Q}_t^\dagger \neq \mathbf{Q}_{2,t}$, even though they have they same form. The first one contains the hyper-parameters of the aggregate series while the second includes the hyperparameters of series 2 only.

The forecasts from the three methods are:

$$\begin{aligned} \hat{y}_{n+1}^{(1)} &= \mathbf{Z}_{n+1}^A \mathbf{a}_{n+1}^{(1)} && \text{Method 1} \\ \hat{y}_{n+1}^{(2)} &= \mathbf{Z}_{n+1}^A \mathbf{a}_{n+1}^{(2)} && \text{Method 2} \\ \hat{y}_{n+1}^{(3)} &= \mathbf{Z}_{n+1}^\dagger \mathbf{a}_{n+1}^{(1)} && \text{Method 3} \end{aligned} \tag{5.2.2}$$

We assume that we have no model identification problem for method 3, but in some cases the above models are not identifiable. The identifiability of an aggregate structural time series model is not straightforward. Even if we have identifiable models for the individual series, we may have an unobserved component models which is not identifiable for the aggregate series if we just add up the components.

We first look at the identifiability conditions for the component series. In the following example we add a non-identifiable model to an identifiable.

Example 5.2.1. For illustration purposes we show that a simple local level+MA(1)

model is not identifiable. Suppose that our model for the j^{th} series is:

$$\begin{aligned}
y_{j,t} &= \mu_{j,t} + z_{j,t} + \epsilon_{j,t} & \epsilon_{j,t} &\sim \text{NID}(0, \sigma_\epsilon^2) \\
\mu_{j,t} &= \mu_{j,t-1} + \eta_{j,t} & \eta_{j,t} &\sim \text{NID}(0, \sigma_\eta^2) \\
z_{j,t} &= \zeta_{j,t} + \theta\zeta_{j,t-1} & \zeta_{j,t} &\sim \text{NID}(0, \sigma_\zeta^2)
\end{aligned} \tag{5.2.3}$$

The reduced form of the model (5.2.3) is an IMA(1,2),

$$\begin{aligned}
\Delta y_t &= \eta_{j,t} + \Delta z_{j,t} + \Delta \epsilon_t \\
&= \eta_{j,t} + \zeta_{j,t} + (\theta - 1)\zeta_{j,t-1} - \theta\zeta_{j,t-2} + \epsilon_{j,t} - \epsilon_{j,t-1}
\end{aligned} \tag{5.2.4}$$

We show the non-identifiability by looking at the autocovariance function of the above model. From (5.2.4), we can calculate the autocovariance function of Δy_t ,

$$\begin{aligned}
\gamma(0) &= \sigma_\eta^2 + \sigma_\zeta^2 + (\theta - 1)^2\sigma_\zeta^2 + \theta\sigma_\zeta^2 + 2\sigma_\epsilon^2 \\
&= \sigma_\eta^2 + [1 + (\theta - 1)^2 + \theta^2]\sigma_\zeta^2 + 2\sigma_\epsilon^2 \\
&= \sigma_\eta^2 + 2(\theta^2 - \theta + 1)\sigma_\zeta^2 + 2\sigma_\epsilon^2 \\
\gamma(1) &= (\theta - 1)\sigma_\zeta^2 - \theta(\theta - 1)\sigma_\zeta^2 - \sigma_\epsilon^2 \\
&= -(\theta - 1)^2\sigma_\zeta^2 - \sigma_\epsilon^2 \\
\gamma(2) &= -\theta\sigma_\zeta^2 \\
\gamma(h) &= 0 \quad \text{for } h \geq 3
\end{aligned}$$

The model is therefore not identifiable since we have three equations but four unknown parameters, i.e σ_ϵ^2 , σ_η^2 , σ_ζ^2 , and θ .

Hotta (1989) shows that for a time series $\{y_t\}$ which has M unobserved components, under very general assumptions, a sufficient condition for identifiability is that $p_i + d_i \geq q_i + 1$, where p_i is the order of the AR polynomial,

d_i is the order of differencing for a non-stationary process, and q_i is the order of the MA polynomial for i^{th} component. If there are no constraints on any of the AR or MA polynomials, then this condition is both necessary and sufficient. For the local level+MA(1) model (5.2.3), the order condition is not satisfied for the MA component, since $p_2 = 0$, $d_2 = 0$, and $q_2 = 1$ so the model is not identifiable. In general, simple MA process in models formulated as in (5.2.3) are not identifiable since the order conditions are not satisfied. On the other hand, we can reformulate the structural models in a way that the MA process have identifiable form. This can be achieved by ignoring the observation error ϵ_t .

In the following example, we show a case where we have two series that satisfy the order condition for identifiability but the aggregate model may be non-identifiable without reformulation.

Example 5.2.2. Suppose we have the following two series that both have AR(1)+error components.

$$\begin{aligned} y_{1,t} &= z_{1,t} + \epsilon_{1,t} \\ z_{1,t} &= \phi_1 z_{1,t-1} + \zeta_{1,t} \quad \zeta_{1,t} \sim \text{NID}(0, \sigma_1^2) \\ y_{2,t} &= z_{2,t} + \epsilon_{2,t} \\ z_{2,t} &= \phi_2 z_{2,t-1} + \zeta_{2,t} \quad \zeta_{2,t} \sim \text{NID}(0, \sigma_2^2) \end{aligned}$$

If we add up the components so that the model for the aggregate series is:

$$\begin{aligned} y_t^A &= z_t^A + \epsilon_t^A \\ z_t^A &= z_{1,t} + z_{2,t} \\ \epsilon_t^A &= \epsilon_{1,t} + \epsilon_{2,t} \end{aligned}$$

Following Granger and Morris (1976) the sum of two AR(1) processes result in an ARMA(p,q), where $p \leq 2$ and $q \leq 1$. The autocovariance function $\gamma(h)$ of the aggregate unobserved component z_t^A as an expression of the parameters ϕ_1 and ϕ_2 are (Harvey 1993):

$$\begin{aligned}\gamma(0) &= \frac{\phi_1^* \theta^* (1 + \phi_2^*) + (1 - \phi_2^*) (\sigma_1^2 + \sigma_2^2)}{(1 - \phi_2^*) \{1 - \phi_1^{*2} \phi_2^* [\phi_1^{*2} + (1 - \phi_2^*) \phi_2^*]\}} \\ \gamma(1) &= \frac{\phi_1^{*2} [\theta^* (1 + \phi_2^*) + (1 - \phi_2^*) (\sigma_1^2 + \sigma_2^2)]}{(1 - \phi_2^*)^2 \{1 - \phi_1^{*2} \phi_2^* [\phi_1^{*2} + (1 - \phi_2^*) \phi_2^*]\}} + \\ &\quad \frac{\theta^* (1 - \phi_2^*) \{1 - \phi_1^{*2} \phi_2^* [\phi_1^{*2} + (1 - \phi_2^*) \phi_2^*]\}}{(1 - \phi_2^*)^2 \{1 - \phi_1^{*2} \phi_2^* [\phi_1^{*2} + (1 - \phi_2^*) \phi_2^*]\}} \\ \gamma(h) &= \phi_1^* \gamma(h-1) + \phi_2^* \gamma(h-2) \quad \text{for } h \geq 2\end{aligned}$$

where,

$$\begin{aligned}\phi_1^* &= \phi_1 + \phi_2 \\ \phi_2^* &= -\phi_1 \phi_2 \\ \theta^* &= \phi_2 \sigma_1^2 + \phi_1 \sigma_2^2\end{aligned}$$

We can easily show that if the following conditions are satisfied,

$$\begin{aligned}\phi_1^* \gamma_{z_t^A}(1) + \phi_2^* \gamma_{z_t^A}(0) &= 0 \\ \phi_2^* \gamma_{z_t^A}(1) &= 0\end{aligned}$$

then z_t^A is a MA(1) process and therefore the model for the aggregate series y_t^A is non-identifiable, unless we ignore the observation error ϵ_t^A .

The above example showed that, in general, it is not sufficient to have identifiable models for the component series to ensure that the aggregate model will be identifiable as well.

For the simple structural time series models, which have a reduced form that follows a MA process, as in the local level model, the local linear trend model and the basic structural model, the aggregation of components is straightforward and always results in identifiable models for the aggregate series. The aggregation of components is not as straightforward when non-seasonal cycles are added in the model which may have reduced forms with AR structure. For the rest of the chapter we deal only with simple structural models with no AR parameters in their reduced form.

5.3 Comparison of forecasts

In the following section we assume that the underlying processes of the individual time series are known and there is no estimation error. We define the following one-step ahead prediction errors:

$$\begin{aligned}v_t^{(1)} &= \mathbf{w}\mathbf{y}_t - \mathbf{Z}_t^A \mathbf{a}_t^{(1)} \\v_t^{(2)} &= \mathbf{w}\mathbf{y}_t - \mathbf{Z}_t^A \mathbf{a}_t^{(2)} \\v_t^{(3)} &= \mathbf{w}\mathbf{y}_t - \mathbf{Z}_t^\dagger \mathbf{a}_t^{(3)}\end{aligned}$$

The variance of the one-step ahead prediction errors is:

$$\begin{aligned}\text{Var}(v_t^{(1)}) &= F_t^{(1)} = \mathbf{Z}_t^A \mathbf{P}_t^{(1)} \mathbf{Z}_t^{A'} + \mathbf{w}\mathbf{H}_t^{(1)} \mathbf{w}' \\ \text{Var}(v_t^{(2)}) &= F_t^{(2)} = \mathbf{Z}_t^A \mathbf{P}_t^{(2)} \mathbf{Z}_t^{A'} + \mathbf{w}\mathbf{H}_t^{(2)} \mathbf{w}' \\ \text{Var}(v_t^{(3)}) &= F_t^{(3)} = \mathbf{Z}_t^\dagger \mathbf{P}_t^\dagger \mathbf{Z}_t^{\dagger'} + \mathbf{H}_t^\dagger\end{aligned}$$

Using the filter derived quantities $F_t^{(1)}$ and $F_t^{(2)}$, we can choose the method which has the smaller variance for the one step ahead prediction error. For all three methods we can evaluate the log-likelihood function using the prediction error decomposition via the Kalman filter. We re-write recursions for the Kalman filter (2.4.11) adding a sub-index j to describe the individual time series:

$$\begin{aligned}
v_{j,t} &= y_{j,t} - \mathbf{Z}_{j,t} \mathbf{a}_{j,t} \\
F_{j,t} &= \mathbf{Z}_{j,t} \mathbf{P}_{j,t} \mathbf{Z}'_{j,t} + H_{j,t} \\
\mathbf{K}_{j,t} &= \mathbf{T}_{j,t} \mathbf{P}_{j,t} \mathbf{Z}'_{j,t} F_{j,t}^{-1} \\
\mathbf{L}_{j,t} &= \mathbf{T}_{j,t} - \mathbf{K}_{j,t} \mathbf{Z}_{j,t} \\
\mathbf{a}_{j,t+1} &= \mathbf{T}_{j,t} \mathbf{a}_{j,t} + \mathbf{K}_{j,t} v_{j,t} \\
\mathbf{P}_{j,t+1} &= \mathbf{T}_{j,t} \mathbf{P}_{j,t} \mathbf{L}'_{j,t} + \mathbf{R}_{j,t} \mathbf{Q}_{j,t} \mathbf{R}'_{j,t}
\end{aligned}$$

Using the prediction error decomposition (2.7.2), we write the log-likelihood function of the three methods as:

$$\log L(\mathbf{wY}_t)^{(k)} = -\frac{n\pi}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^n (\log F_t^{(k)} + v_t^{(k)} F_t^{(k)-1} v_t^{(k)})$$

for $k = 1, 2, 3$. Using an information criterion, such as AIC, we compare the three different methods. We then select the method which has smaller AIC. The use of AIC opens up a simple way which can be used to chose between methods 1 and 3. As we said at the beginning, the comparison depends on the data generating process, so an obvious choice is to use an information criterion such as AIC.

5.4 Estimated coefficients

Up to now, we assumed that the underlying process is known and no coefficients need to be estimated. In practice, the coefficients need to be estimated and there is an uncertainty on the value of the estimates. The coefficients of all models are assumed to be estimated with maximum likelihood using realizations from processes independent of these used for forecasting and identical stochastic structures. The asymptotic conditional variance of the one step ahead prediction error is (Ansley and Kohn 1986),

$$\text{Var}(v_t^{(k)} | \mathbf{Y}_t) = F_t^{(k)} + \frac{1}{n} \frac{\partial v_t^{(k)}}{\partial \boldsymbol{\psi}} \text{Var}(\boldsymbol{\psi}) \frac{\partial v_t^{(k)}}{\partial \boldsymbol{\psi}'} \quad (5.4.1)$$

where $\boldsymbol{\psi}$ is the vector of estimated coefficients and $k = 1, 2, 3$. In the case of methods 1 and 2 the asymptotic distribution of the conditional variance of the aggregated series is,

$$\text{Var}(v_t^{(k)} | \mathbf{Y}_t) = \mathbf{Z}_t^A \mathbf{P}_t^k \mathbf{Z}_t^{A'} + \mathbf{H}_t^k + \frac{1}{n} \frac{\partial v_t^{(k)}}{\partial \boldsymbol{\psi}} \text{Var}(\boldsymbol{\psi}) \frac{\partial v_t^{(k)}}{\partial \boldsymbol{\psi}'}$$

For large samples, the second term is well approximated by zero but in small samples, the estimation variability can be substantial. The asymptotic distribution will be (Harvey 1989, p.211),

$$\text{Var}(\boldsymbol{\psi}) = 2\mathbf{I}(\boldsymbol{\psi})^{-1}$$

where $\mathbf{I}(\boldsymbol{\psi})$ is the information matrix. In the time domain the ij^{th} element of the information matrix is given by (Harvey 1989, p.142):

$$I_{ij}(\boldsymbol{\psi}) = \frac{1}{2} \sum_{t=1}^n \left[\text{tr} \left[F_t^{(k)-1} \frac{\partial F_t^{(k)}}{\partial \boldsymbol{\psi}_i} F_t^{(k)-1} \frac{\partial F_t^{(k)}}{\partial \boldsymbol{\psi}_j} \right] \right] + \text{E} \left[\sum_{t=1}^n \frac{\partial v_t^{(k)'}}{\partial \boldsymbol{\psi}_i} F_t^{(k)-1} \frac{\partial v_t^{(k)}}{\partial \boldsymbol{\psi}_j} \right]$$

Dropping the expectation from the second term, we have an expression which is asymptotically equivalent and in most case it is easier to evaluate.

Alternatively, the information matrix can be written in the frequency domain form as (Harvey 1989, p.196):

$$\mathbf{I}(\boldsymbol{\psi}) = \frac{1}{2} \sum_{j=1}^{n-1} \frac{1}{g_j^2} \frac{\partial g_j}{\partial \boldsymbol{\psi}} \frac{\partial g_j}{\partial \boldsymbol{\psi}'}$$

where g_j depend on the autocovariance generating function of the model. The asymptotic variance of hyperparameters is then directly dependent on the model we chose. For example for the basic structural model is,

$$g_j = 2(1 - \cos \lambda_j s) \sigma_\eta^2 + [s + 2 \sum_{h=1}^{s-1} (s-h) \cos \lambda_j h] \sigma_\zeta^2$$

$$(6 - 8 \cos \lambda_j + 2 \cos 2\lambda_j) \sigma_\omega^2 + 4(1 - \cos \lambda_j)(1 - \cos \lambda_j s) \sigma_\epsilon^2$$

$$\frac{\partial g_j}{\partial \boldsymbol{\psi}} = \begin{bmatrix} 2(1 - \cos \lambda_j s) \\ s + 2 \sum_{h=1}^{s-1} (s-h) \cos \lambda_j h \\ 6 - 8 \cos \lambda_j + 2 \cos 2\lambda_j \\ 4(1 - \cos \lambda_j)(1 - \cos \lambda_j s) \end{bmatrix}$$

The derivatives of F_t and v_t may be evaluated numerically or analytically. Analytical evaluation requires n additional sets of recursions to run in parallel with the Kalman filter. For simplicity we drop the superindex k . Taking

partial derivatives we can have the following recursions:

$$\begin{aligned}
\frac{\partial v_t}{\partial \psi_i} &= -\frac{\partial \mathbf{Z}_t}{\partial \psi_i} \mathbf{a}_t - \mathbf{Z}_t \frac{\partial \mathbf{a}_t}{\partial \psi_i} \\
\frac{\partial \mathbf{a}_{t+1}}{\partial \psi_i} &= \frac{\partial \mathbf{T}_t}{\partial \psi_i} \mathbf{a}_t + \mathbf{T}_t \frac{\partial \mathbf{a}_t}{\partial \psi_i} + \frac{\partial \mathbf{K}_t}{\partial \psi_i} v_t + \mathbf{K}_t \frac{\partial v_t}{\partial \psi_i} \\
\frac{\partial \mathbf{K}_t}{\partial \psi_i} &= \frac{\partial \mathbf{T}_t}{\partial \psi_i} \mathbf{P}_t \mathbf{Z}_t' F_t^{-1} + \mathbf{T}_t \frac{\partial \mathbf{P}_t}{\partial \psi_i} \mathbf{Z}_t' F_t^{-1} + \mathbf{T}_t \mathbf{P}_t \frac{\partial \mathbf{Z}_t'}{\partial \psi_i} F_t^{-1} + \mathbf{T}_t \mathbf{P}_t \mathbf{Z}_t' \frac{\partial F_t^{-1}}{\partial \psi_i} \\
\frac{\partial F_t}{\partial \psi_i} &= \frac{\partial \mathbf{Z}_t}{\partial \psi_i} \mathbf{P}_t \mathbf{Z}_t' + \mathbf{Z}_t \frac{\partial \mathbf{P}_t}{\partial \psi_i} \mathbf{Z}_t' + \mathbf{Z}_t \mathbf{P}_t \frac{\partial \mathbf{Z}_t'}{\partial \psi_i} + \frac{\partial H_t}{\partial \psi_i} \\
\frac{\partial F_t^{-1}}{\partial \psi_i} &= -F_t^{-1} \frac{\partial F_t}{\partial \psi_i} F_t^{-1} \\
\frac{\partial \mathbf{L}_t}{\partial \psi_i} &= \frac{\partial \mathbf{T}_t}{\partial \psi_i} - \frac{\partial \mathbf{K}_t}{\partial \psi_i} \mathbf{Z}_t - \mathbf{K}_t \frac{\partial \mathbf{Z}_t}{\partial \psi_i} \\
\frac{\partial \mathbf{P}_{t+1}}{\partial \psi_i} &= \frac{\partial \mathbf{T}_t}{\partial \psi_i} \mathbf{P}_t \mathbf{L}_t' + \mathbf{T}_t \frac{\partial \mathbf{P}_t}{\partial \psi_i} \mathbf{L}_t' + \mathbf{T}_t \mathbf{P}_t \frac{\partial \mathbf{L}_t}{\partial \psi_i} + \mathbf{R}_t \frac{\partial \mathbf{Q}_t}{\partial \psi_i} \mathbf{R}_t'
\end{aligned} \tag{5.4.2}$$

If \mathbf{a}_0 and \mathbf{P}_0 are independent of ψ as in the case of diffuse prior, then $\frac{\partial \mathbf{a}_t}{\partial \psi_i} = 0$ and $\frac{\partial \mathbf{P}_t}{\partial \psi_i} = 0$. Harvey (1989, p.143) provides some of the recursions in (5.4.2) for a different formulation of the Kalman filter. Here we derive the complete set of recursions needed for the calculation of the asymptotic variance (5.4.1) using the formulation of the Kalman filter seen in (2.4.11).

In the case of the exact initial Kalman filter (2.6.1), we also derive recursions

in a similar way,

$$\begin{aligned}
\frac{\partial \mathbf{a}_{t+1}}{\partial \psi_i} &= \frac{\partial \mathbf{T}_t}{\partial \psi_i} \mathbf{a}_t + \mathbf{T}_t \frac{\partial \mathbf{a}_t}{\partial \psi_i} + \frac{\partial \mathbf{K}_{*,t}}{\partial \psi_i} v_t + \mathbf{K}_{*,t} \frac{\partial v_t}{\partial \psi_i} \\
\frac{\partial \mathbf{P}_{*,t+1}}{\partial \psi_i} &= \frac{\partial \mathbf{T}_t}{\partial \psi_i} \mathbf{P}_{*,t} \mathbf{T}'_t + \mathbf{T}_t \frac{\partial \mathbf{P}_{*,t}}{\partial \psi_i} \mathbf{T}'_t + \mathbf{T}_t \mathbf{P}_{*,t} \frac{\partial \mathbf{T}'_t}{\partial \psi_i} - \frac{\partial \mathbf{C}_{*,t}}{\partial \psi_i} + \mathbf{R}_t \frac{\partial \mathbf{Q}_t}{\partial \psi_i} \mathbf{R}'_t \\
\frac{\partial \mathbf{P}_{\infty,t+1}}{\partial \psi_i} &= \frac{\partial \mathbf{T}_t}{\partial \psi_i} \mathbf{P}_{\infty,t} \mathbf{T}'_t + \mathbf{T}_t \frac{\partial \mathbf{P}_{\infty,t}}{\partial \psi_i} \mathbf{T}'_t + \mathbf{T}_t \mathbf{P}_{\infty,t} \frac{\partial \mathbf{T}'_t}{\partial \psi_i} - \frac{\partial \mathbf{C}_{\infty,t}}{\partial \psi_i} \\
\frac{\partial \mathbf{K}_{*,t}}{\partial \psi_i} &= \frac{\partial \mathbf{M}_{*,t}}{\partial \psi_i} \mathbf{F}_{*,t}^- + \mathbf{M}_{*,t} \frac{\partial \mathbf{F}_{*,t}^-}{\partial \psi_i} + \frac{\partial \mathbf{M}_{\infty,t}}{\partial \psi_i} \mathbf{F}_{\infty,t}^- + \mathbf{M}_{\infty,t} \frac{\partial \mathbf{F}_{\infty,t}^-}{\partial \psi_i} \\
\frac{\partial \mathbf{F}_{*,t}^-}{\partial \psi_i} &= 2J_{1,t} \frac{\partial J_{1,t}}{\partial \psi_i} \\
\frac{\partial \mathbf{F}_{\infty,t}^-}{\partial \psi_i} &= 2J_{2,t} \frac{\partial J_{2,t}}{\partial \psi_i} \\
\frac{\partial \mathbf{M}_{*,t}}{\partial \psi_i} &= \frac{\partial \mathbf{T}_t}{\partial \psi_i} \mathbf{P}_{*,t} \mathbf{Z}'_t + \mathbf{T}_t \frac{\partial \mathbf{P}_{*,t}}{\partial \psi_i} \mathbf{Z}'_t + \mathbf{T}_t \mathbf{P}_{*,t} \frac{\partial \mathbf{Z}'_t}{\partial \psi_i} \\
\frac{\partial \mathbf{M}_{\infty,t}}{\partial \psi_i} &= \frac{\partial \mathbf{T}_t}{\partial \psi_i} \mathbf{P}_{\infty,t} \mathbf{Z}'_t + \mathbf{T}_t \frac{\partial \mathbf{P}_{\infty,t}}{\partial \psi_i} \mathbf{Z}'_t + \mathbf{T}_t \mathbf{P}_{\infty,t} \frac{\partial \mathbf{Z}'_t}{\partial \psi_i} \\
\frac{\partial \mathbf{C}_{*,t}}{\partial \psi_i} &= \frac{\partial \mathbf{M}_{*,t}}{\partial \psi_i} \mathbf{K}_{*,t} + \mathbf{M}_{*,t} \frac{\partial \mathbf{K}_{*,t}}{\partial \psi_i} \\
&\quad + \left(\frac{\partial \mathbf{M}_{\infty,t}}{\partial \psi_i} \mathbf{F}_{\infty,t}^- + \mathbf{M}_{\infty,t} \frac{\partial \mathbf{F}_{\infty,t}^-}{\partial \psi_i} \right) (\mathbf{M}_{*,t} - \mathbf{M}_{\infty,t} \mathbf{F}_{\infty,t}^- \mathbf{F}_{*,t})' \\
&\quad + \mathbf{M}_{\infty,t} \mathbf{F}_{\infty,t}^- \left(\frac{\partial \mathbf{M}_{*,t}}{\partial \psi_i} - \frac{\partial \mathbf{M}_{\infty,t}}{\partial \psi_i} \mathbf{F}_{\infty,t}^- \mathbf{F}_{*,t} - \mathbf{M}_{\infty,t} \frac{\partial \mathbf{F}_{\infty,t}^-}{\partial \psi_i} \mathbf{F}_{*,t} \right. \\
&\quad \quad \left. - \mathbf{M}_{\infty,t} \mathbf{F}_{\infty,t}^- \frac{\partial \mathbf{F}_{*,t}^-}{\partial \psi_i} \right)' \\
\frac{\partial \mathbf{C}_{\infty,t}}{\partial \psi_i} &= \frac{\partial \mathbf{M}_{\infty,t}}{\partial \psi_i} \mathbf{F}_{\infty,t}^- \mathbf{F}_{*,t} + \mathbf{M}_{\infty,t} \frac{\partial \mathbf{F}_{\infty,t}^-}{\partial \psi_i} \mathbf{F}_{*,t} + \mathbf{M}_{\infty,t} \mathbf{F}_{\infty,t}^- \frac{\partial \mathbf{F}_{*,t}^-}{\partial \psi_i}
\end{aligned} \tag{5.4.3}$$

Even though $\frac{\partial J_{1,t}}{\partial \psi_i}$ and $\frac{\partial J_{2,t}}{\partial \psi_i}$ are time dependent, in practice are quite easy to estimate using simple matrix calculus. When $\mathbf{P}_{\infty,t} = 0$ we can use the first set of recursion to evaluate the necessary partial derivative.

Having defined all the necessary quantities, using either (5.4.2) or (5.4.3), we

can calculate the variance of the one step ahead prediction error for the case of estimated coefficients, given in (5.4.1). We can follow the same idea we introduced in §5.3 and compare the three estimation methods (5.2.2). We then choose the one which will give the smallest conditional variance for the one step ahead prediction error of the aggregate series.

5.5 Application

We apply the methodology for selecting the best forecasting method in the UK Motor Vehicle Production dataset (Figure 5.1). The total car manufacturing is disaggregated into production for the home market and production for exports. They belong to a set of short-term output indicators used to determine economic policy. All three series have equal importance for different reasons. The total production is an indicator for the state of manufacturing in UK, while the production for home can also be used as an indicator for the aggregate demand in the British economy. The production for exports is used to determine the effect of exchange rates in the competitiveness of the UK economy. Given their equal importance, it is interesting to decide what is the best way to forecast each of three series. Each one can be forecast indirectly using either method 1 or 2, or directly using method 3. For example, a forecast for the home market production can be estimated directly or as the difference between the total and export production forecasts. We applied the three methods in each series. The results are shown in Table

For all three series, method 3, provides the smallest variance of the one-step ahead prediction error. We also give the variance adjusted for the uncertainty on the value of the estimates. The asymptotic conditional variance when the coefficients are unknown is, as expected, larger than if we assume known parameters. The largest contribution due to parameter uncertainty is in method 1, for home and export production, and in method 2 for the total production.

We see that in terms of AIC, total production is better modeled directly while export production is better modeled indirectly. For the home production, methods 1 and 3 behave similarly. In some cases, analysts are interested in having consistency in the aggregation of forecasts, so that the forecasts for the home and export production add up to the forecast for the total production. Applying our methodology gives the best way to provide consistently aggregated forecasts. Total and home production would be modeled directly, while forecast for the export series would be a linear combination of the forecasts of the two other series. The export and the total series are very similar, especially in recent years. The results of this analysis show that for forecasting purposes, there is no significant information loss for the export series if we use an indirect method to forecast it; all the information needed is included in the total series. The opposite cannot be said for the total series; there is information loss if use an indirect method for the total series. The same methodology can be followed for seasonal adjustment purposes. ONS published seasonally adjusted series for total production by adding up the seasonally adjusted estimates for the home and the export

Table 5.1: Comparison of direct and indirect methods

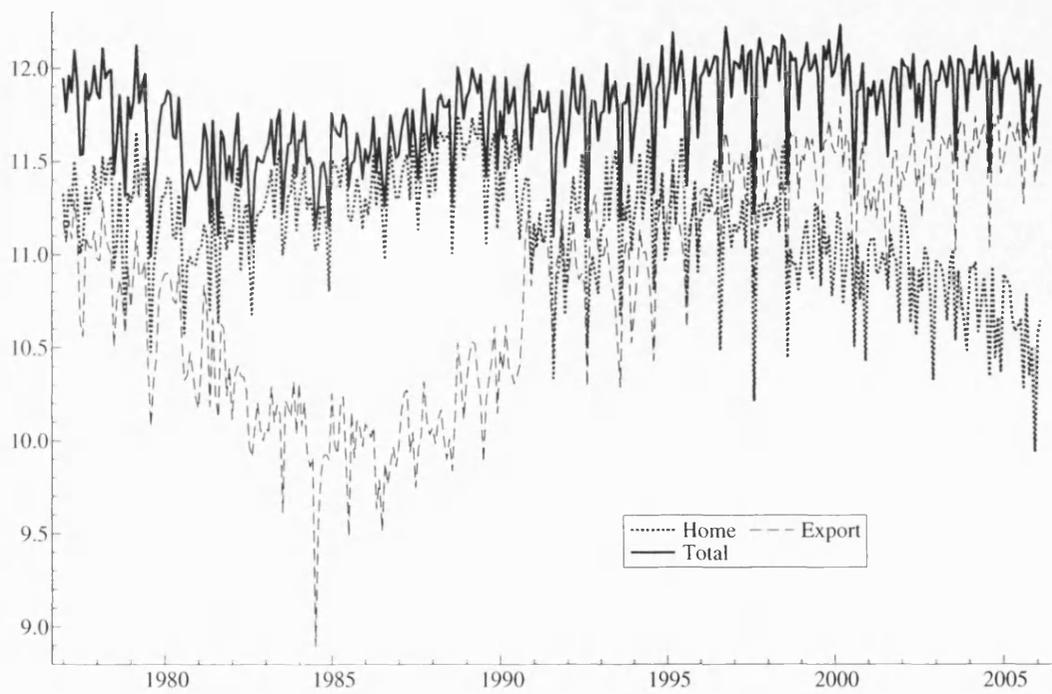
	Home	Export	Total
$AIC^{(1)}$	-4.9820	-5.1120	-3.1088
$AIC^{(2)}$	-4.4892	-4.6188	-3.0618
$AIC^{(3)}$	-4.9491	-5.0600	-6.0111
$F_n^{(1)}$	0.041919	0.042732	0.052737
$F_n^{(2)}$	0.062811	0.062805	0.071530
$F_n^{(3)}$	0.027367	0.026484	0.016993
Adj. $F_n^{(1)}$	0.042442	0.043645	0.052739
Adj. $F_n^{(2)}$	0.062816	0.062814	0.071897
Adj. $F_n^{(3)}$	0.027368	0.026384	0.017044

series. Following a methodological review in 2004 (Tripodis 2005), total and home production and seasonally adjusted directly, while the export series is derived indirectly.

5.6 Conclusions

In many areas of economic life, we need forecasts of aggregated variables. Practitioners have to face the question of whether to forecast the components of a dataset and add up the forecasts (indirect method), or to forecast the aggregate series separately (direct method). This chapter investigated the differences between these methods. We consider the similarities and differences within the structural time series framework. We point out some

Figure 5.1: Motor Vehicle Production in the UK



problems in the identifiability for the model of the aggregate series. We also suggest a methodology to take an appropriate decision for each series. We start from the case where the underlying process of the series is known and we base our comparison of the methods on the variance of the one step ahead forecast error. This can be used either directly or through some information criterion to provide the best method for each case. We also look at the more common case where the underlying process is unknown and the parameters of the model need to be estimated. In this case, we provide recursion which will give the variance of the one-step ahead prediction error and follow the same methodology as in the case where the parameters are known. We applied this methodology in the UK motor vehicle production dataset. We showed that it is better to model directly the total and home series, and derive the forecasts for the export series by subtracting the home from the total production forecasts.

Chapter 6

Conclusion and further research

This thesis concentrates on two particular aspects of time series: seasonal heteroscedasticity and aggregation. Economic time series often have one season with different variance from the others. We show that single season heteroscedasticity can be represented using periodic seasonal variance or periodic irregular variance; these approaches differ markedly in their periodic covariance structure. We suggest that periodic seasonal variance is more appropriate for modelling economic time series. In economic series, higher variability in one season is usually a feature of the seasonal component; economic agents have knowledge of seasonality and are able to counteract higher variability in one season by adjusting their behaviour in all other seasons. For example, factory owners may compensate for low production in one month by increased production in subsequent months. Higher variability in a single season may also result from an exogenous effect. In the case of natural events, such as river flow, the exogenous effect is rainfall which is not balanced across the year. For series where there is no mechanism to balance

higher variability in one season across the other seasons, a periodic irregular model may be more appropriate. We show that a likelihood ratio test is effective in detecting single season heteroscedasticity. Our test also allows us to distinguish between cases with periodic variance in the irregular component and those where periodic seasonal variance provides a better model.

We extended the case of single season heteroscedasticity to general periodic process, where the coefficients change with the season. The most widely used framework is the periodic autoregression (PAR). We advocate representation of periodic processes in the structural time series framework. We established that, periodic structural models have a vector integrated moving average representation. Although vector moving average and vector integrated moving average models have been investigated as part of VARIMA class of models, they are little used in practice. We have shown that a class of models with VIMA correlation structure provide parsimonious models for univariate series with periodic second moments. Moreover, in practical applications, the structural framework provides greater insight into the nature of the series than PAR models. PSTSM relate the seasonality in the autocovariance function to specific unobserved components. In comparison, PAR models parameters are not readily interpretable. We compare the forecasting accuracy of PSTSM with PAR models. PSTSM produced better forecasts both within and out of sample for the majority of the series concerned. We conclude that structural time series models are a natural framework for modelling periodic processes both in terms of interpretability and forecasting accuracy.

We also consider the question whether to forecast the components of a dataset

and add up the forecasts, or forecast the aggregate series separately. Many economic time series can be broken down into components, such as regions or different types of economic activity. We investigated the differences between the different methods. We suggest a methodology to take an appropriate decision for each series. We start from the case where the underlying process of the series is known and we base our comparison of the methods on the variance of the one step ahead forecast error. This can be used either directly or through some information criterion, such as AIC, to provide the best method for each case. We also look at the more common case where the underlying process is unknown and the parameters of the model need to be estimated. In this case, we provide recursion which gives the variance of the one-step ahead prediction error and follow the same methodology as in the case where the parameters are known.

An interesting direction for further research is to combine the issue of seasonal heteroscedasticity within a large dataset with different component and see the interactions across different time series. We need to investigate further issues of seasonal co-integration that were not discussed in this thesis. In many datasets, the seasonal component has common cause and one can envisage situations where seasonal components are co-integrated. Extending the work by Hylleberg, Engle, Granger, and Yoo (1990), Harvey and Koopman (1997) propose common seasonal models within the structural time series framework. Bringing together all these aspects of seasonal time series, would provide considerable advantages in terms of forecasting of time series and also in terms of identification of seasonal component for the purpose of

removing it through the process of seasonal adjustment.

Appendix A

A.1 Givens Rotations algorithm

We use the algorithm provide by Golum and van Loan (1996). Given scalars a and b , this functions computes $c = \cos(\theta)$ and $s = \sin(\theta)$ so

$$\begin{bmatrix} c & s \\ -s & c \end{bmatrix}^T \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} r \\ 0 \end{bmatrix}$$

if $b = 0$

$$c = 1; s = 0$$

else

if $|b| > |a|$

$$\tau = -\frac{a}{b}$$

$$s = \frac{1}{\sqrt{1+\tau^2}}$$

$$c = sr$$

else

$$\tau = -\frac{a}{b}$$

$$c = \frac{1}{\sqrt{1+r^2}}$$

$$s = cr$$

end

end The following function written in Ox (Doornik 1998) applies repeated Givens rotation to the input matrix **mU** to create a lower triangular matrix.

```
Givens()  
{  
  decl mG,c,s,x1,x2;  
  decl i,j;  
  decl cr=rows(mU);  
  decl ccol=columns(mU);  
  decl k;  
  decl tau,t1,t2;  
  for(k=ccol-1;k>0;k--)  
  {  
    for(i=0;i<cr;i++)  
    {  
      if(i==k)  
        return -1;  
      else  
      {  
        t1=mU[i][i];
```

```

t2=mU[i][k];
    if(t2==0)
        {
            c=1;
            s=0;
        }
    else
        {
            if(sqrt(t2^2)>sqrt(t1^2))
                {
                    tau=-t1/t2;
                    s=1/sqrt(1+tau^2);
                    c=s*tau;
                }
            else
                {
                    tau=-t2/t1;
                    c=1/sqrt(1+tau^2);
                    s=c*tau;
                }
        }
mU[i][i]=c*t1-s*t2;
mU[i][k]=s*t1+c*t2;

```

```

        }
    }
}

```

A.2 Square Kalman Filter code

The following function written in Ox gives the square root Kalman filter

```

SqFilter()
{
    decl i;
    decl mZ=mTZ[cS] [];          // system matrix Z
    decl mT=mTZ[:cS-1] [];      // system matrix T
    decl mH,mL,mC;
    decl mRQ=mHG HG[0:cS-1][0:cS-1]; //variance of state noise
    decl mState=zeros(cS,cDataLength);
    decl mCState=zeros(cS,1);
    decl mU2=zeros(cS,1);
    decl mU3=zeros(cS,cS);
    decl mU1=0;
    mState[] [0]=0;
    vFInv=zeros(1,cDataLength); //cDataLength: no of observations

```

```

vVlow=zeros(1,cDataLength);
vVlow[0][0]=mData[0][0]; //mData is input matrix with original
                        // data
if(cSeasType==CMP_SEAS_HS) //Harrison-Stevens seasonality
{
  decl result=eigensym(mRQ,&mL,&mC);
  if(result==1) return -1;
  else
  {
    for(i=0;i<cS;i++)
    {
      if(mL[i]<0) mL[i]=0;
      else mL[i]=mL[i]^0.5;
    }
    mL=setdiagonal(zeros(cS,cS),mL);
    mU=mC*mL;
    Givens();
    mRQ=mU;
    mX=mX';
  }
}
else mRQ=choleski(mRQ);
mU=zeros(cSy,2*cS+1);

```

```

if(cSeasType==CMP_SEAS_HS) mZ[] [cS-cSeasPer:cS-1]=mX[0] [];
mU[0] [:cS-1]=mZ*mP;
mU[1:cS] [0:cS-1]=mT*mP;
if(cSeasType==CMP_SEAS_DUMMY) mU[0] [cS]=mX[0] []^0.5;
else mU[0] [cS]=mX[0] [cSeasPer]^0.5;
mU[1:] [cSy:]=mRQ;
Givens();
vFInv[] [0]=mU[0] [0]^(-2);
mState[] [1]=mT*mState[] [0]+mU2*mU1*vVlow[] [0];
for(i=1;i<cDataLength-1;i++)
{
mU=zeros(cSy,2*cS+1);
if(cSeasType==CMP_SEAS_HS) mZ[] [cS-cSeasPer:cS-1]=mX[i] [];
vVlow[] [i]=mData[] [i]-mZ*mState[] [i];
mU[0] [:cS-1]=mZ*mP;
mU[1:cS] [0:cS-1]=mT*mP;
if(cSeasType==CMP_SEAS_DUMMY) mU[0] [cS]=mX[i] []^0.5;
else mU[0] [cS]=mX[i] [cSeasPer]^0.5;
mU[1:] [cSy:]=mRQ;
Givens();
vFInv[] [i]=mU[0] [0]^(-2);
mU1=1/mU[0] [0];
mU2=mU[1:] [0];

```

```

mU3=mU[1:][1:cS];
mU3=setupper(mU3,0);
mP=mU3;
mState[][i+1]=mT*mState[][i]+mU2*mU1*vVlow[][i];
}
if(cSeasType==CMP_SEAS_HS)mZ[][cS-cSeasPer:cS-1]=mX[i][];
vVlow[][cDataLength-1]=mData[][cDataLength-1]-mZ*mState[][cDataLength-1];
mU[0][:cS-1]=mZ*mP;
mU[1:cS][0:cS-1]=mT*mP;
if(cSeasType==CMP_SEAS_DUMMY) mU[0][cS]=mX[i][]^0.5;
else mU[0][cS]=mX[i][cSeasPer]^0.5;
mU[1:][cSy:]=mRQ;
Givens();
vFInv[][cDataLength-1]=mU[0][0]^(-2);
mX=mX';
}

```

Bibliography

- Akaike, H. (1974). A new look at the statistical model identification. *Annals of the Institute of Statistical Mathematics* 21, 243–247.
- Anderson, B. and J. Moore (1979). *Optimal Filtering*. Englewood Cliffs: Prentice Hall.
- Anderson, P., M. Meerschaert, and A. Vecchia (1999). Innovations algorithm for periodically stationary time series. *Stochastic Processes and their Applications* 3, 149–169.
- Anderson, T. (1984). *An Introduction to Multivariate Statistical Analysis*. New York: John Wiley and Sons.
- Ansley, C. F. and R. Kohn (1985). Estimation, filtering and smoothing in state space models with incompletely specified initial conditions. *Annals of Statistics* 13, 1286–1316.
- Ansley, C. F. and R. Kohn (1986). Prediction mean square error for state-space models with estimated parameters. *Biometrika* 73, 467–474.
- Basawa, I. and R. Lund (2001). Large sample properties of parameter estimates for periodic arma models. *Journal of Time Series Analysis* 22,

651–663.

- Bell, W. R. and S. C. Hillmer (1984). Issues involved with the seasonal adjustment of economic time series. *Journal of Business and Economic Statistics* 2, 291–320.
- Bowman, K. O. and L. R. Shenton (1975). Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2 . *Biometrika* 62, 243–250.
- Burrige, P. and K. F. Wallis (1988). Seasonal adjustment and Kalman filtering: extension to periodic variances. *Journal of Forecasting* 9, 109–118.
- de Jong, P. (1988). A cross validation filter for time series models. *Biometrika* 75, 594–600.
- de Jong, P. (1989). Smoothing and interpolation with the state space model. *Journal of the American Statistical Association* 84, 1085–1088.
- de Jong, P. (1991). The diffuse Kalman filter. *Annals of Statistics* 19, 1073–1083.
- Doornik, J. (1998). *Object-Oriented Matrix Programming using Ox 2.0*. Timberlake Consultants Press.
- Durbin, J. and S. Koopman (2001). *Time Series Analysis by State Space Methods*. Oxford University Press.
- Engle, R. (1984). Wald, likelihood ratio, and Lagrange multiplier tests in econometrics. In Z. Griliches and M. Intriligator (Eds.), *Handbook of Econometrics*, Volume 2. Amsterdam: North Holland.

- Findley, D. (2005). Some recent developments and directions in seasonal adjustment. *Journal of Official Statistics* 21, 343–365.
- Findley, D., B. Monsell, M. Otto, W. Bell, and M. Pugh (1998). New capabilities and methods of the X-12-ARIMA seasonal adjustment program. *Journal of Business and Economic Statistics* 16, 127–177.
- Fletcher, R. (1987). *Practical Methods of Optimisation*. New York: John Wiley.
- Franses, P. H. (1996). *Periodicity and Stochastic Trends in Economic Time Series*. Oxford University Press.
- Ghysels, E. and D. Osborn (2001). *The Econometric Analysis of Seasonal Time Series*. Cambridge University Press.
- Giacomini, R. and C. Granger (2004). Aggregation of space-time processes. *Journal of Econometrics* 118, 7–26.
- Gladyshev, E. G. (1961). Periodically correlated random sequences. *Soviet Mathematics* 2, 385–388.
- Goldfeld, S. and R. Quandt (1965). Some tests for homoscedasticity. *Journal of the American Statistical Association* 60, 539–547.
- Golum, G. and C. F. van Loan (1996). *Matrix Computations*. The John Hopkins University Press.
- Gomez, V. and A. Maravall (1996). Programs TRAMO and SEATS. Technical Report 9628, Bank of Spain.
- Granger, C. and M. Morris (1976). Time series modelling and interpretation. *Journal Royal Statistical Society A* 139, 246–256.

- Hamilton, J. (1994). *Time Series Analysis*. Princeton University Press.
- Hannan, E. J., R. D. Terrell, and N. E. Tuckwell (1970). The seasonal adjustment of economic time series. *International Economic Review* 11, 24–52.
- Harrison, P. J. and C. F. Stevens (1976). Bayesian forecasting. *Journal of the Royal Statistical Society, Series B* 38, 207–247.
- Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press.
- Harvey, A. C. (1993). *Time Series Models*. Hemel Hempstead: Harvester Wheatsheaf.
- Harvey, A. C. and S. J. Koopman (1997). Multivariate structural time series models. In *Systematic Dynamics in Economic and Financial models*, pp. 269–298. Chichester: John Wiley and Sons.
- Hipel, K. and I. McLeod (1994). *Time Series Modelling of Water Resources and Environmental Systems*. Amsterdam:Elsevier.
- Hotta, L. (1989). Identification of unobserved component models. *Journal of Time Series Analysis* 10, 259–270.
- Hotta, L. and K. Vasconcellos (1999). Aggregation and disaggregation of structural time series models. *Journal of Time Series Analysis* 20, 155–171.
- Hurd, H. and N. Gerr (1991). Graphical methods for determining the presence of periodic correlation. *Journal of Time Series* 12, 337–350.
- Hylleberg, S. (1992). *Modelling seasonality*. Oxford University Press.

- Hylleberg, S., R. Engle, C. Granger, and B. Yoo (1990). Seasonal integration and cointegration. *Journal of Econometrics* 44, 215–238.
- Jones, R. and W. Brelsford (1967). Time series with periodic structure. *Biometrika* 54, 403–408.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering, Transactions ASMA, Series D* 82, 35–45.
- Kohn, R. and C. F. Ansley (1989). A fast algorithm for signal extraction, influence and cross-validation. *Biometrika* 76, 65–79.
- Koopman, S. J. (1993). Disturbance smoother for state space models. *Biometrika* 80, 117–126.
- Koopman, S. J. (1997). Exact initial Kalman filtering and smoothing for non-stationary time series models. *Journal of the American Statistical Association* 92, 1630–1638.
- Koopman, S. J. and N. Shephard (1992). Exact score for time series models in state space form. *Biometrika* 79, 823–826.
- Koopman, S. J., N. Shephard, and J. A. Doornik (1998). Statistical algorithms for models in state space using SsfPack 2.2. *Econometric Journal* 1, 1–55.
- Ljung, G. M. and G. E. P. Box (1978). On a measure of lack of fit in time series models. *Biometrika* 66, 67–72.
- Lund, R. and I. Basawa (2000). Recursive prediction and likelihood evaluation for periodic arma models. *Journal of Time Series Analysis* 21,

75–93.

- Lütkepohl, H. (1987). *Forecasting aggregated vector ARMA processes*. Springer-Verlag.
- Magnus, J. R. and H. Neudecker (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. New York: Wiley.
- Miron, J. A. (2001). *The Economics of Seasonal Cycles*. The MIT Press.
- Morf, J. F. and T. Kailath (1975). Square root algorithms for least square estimation. *IEEE Transactions on Automatic Control* 20, 487–497.
- Novales, A. and R. F. de Frutto (1997). Forecasting with periodic models: A comparison with time invariant coefficient models. *International Journal of Forecasting* 13, 393–405.
- Osborn, D. and J. Smith (1988). Seasonality and habit persistence in a life cycle model of consumption. *Journal of Applied Econometrics* 3, 255–266.
- Osborn, D. and J. Smith (1989). The performance of periodic autoregressive models in forecasting seasonal U.K. consumption. *Journal of Business Economic and Statistics* 7, 117–127.
- Pagano, M. (1978). On periodic and multiple autoregressions. *Annals of Statistics* 6, 1310–1317.
- Parzen, E. and M. Pagano (1979). An approach to modeling seasonally stationary time series. *Journal of Econometrics* 9, 137–153.
- Priestley, M. (1981). *Spectral Analysis and Time Series*. Academic Press.

- Proietti, T. (1998). Seasonal heteroscedasticity and trends. *Journal of Forecasting* 17, 1–17.
- Proietti, T. (2000). Comparing seasonal components for structural time series models. *International Journal of Forecasting* 16, 247–260.
- Rose, D. (1977). Forecasting aggregates of independent arima processes. *Journal of Econometrics* 5, 323–345.
- Shao, Q. and R. Lund (2004). Computation and characterization of autocorrelations and partial autocorrelations in periodic arma models. *Journal of Time Series Analysis* 25, 359–372.
- Tiao, G. and I. Guttman (1980). Forecasting contemporaneous aggregates of multiple time series. *Journal of Econometrics* 12, 219–230.
- Tripodis, Y. (2005). Methodological improvements to the seasonal adjustment of the motor vehicle production index. *Economic Trends* 619, 48–53.
- Tripodis, Y. and J. Penzer (2006). Single season heteroscedasticity in time series. *Journal of Forecasting* (forthcoming).
- Troutman, B. (1979). Some results in periodic autoregression. *Biometrika* 66, 219–228.
- Vecchia, A. and R. Ballerini (1991). Testing for periodic autocorrelations in seasonal time series data. *Biometrika* 78, 53–63.
- Wei, W. and J. Abraham (1981). Forecasting contemporaneous time series aggregates. *Communications in Statistics - Theory and Methods* A10, 1335–1344.

White, H. (1980). A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. *Econometrica* 48, 817–838.

Whittle, P. (1963). On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix. *Biometrika* 50, 129–134.