Filtering Non-Linear State Space Models: Methods and Economic Applications
ISBN 978 90 3610 169 1

Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul

This book is no. 474 of the Tinbergen Institute Research Series, established through cooperation between Thela Thesis and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.
Filtering Non-Linear State Space Models: Methods and Economic Applications

ACADEMISCH PROEFSCHRIFT

der verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. L.M. Bouter,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Economische Wetenschappen en Bedrijfskunde
op dinsdag 4 mei 2010 om 13.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door
Kai Ming Lee
geboren te Hong Kong, Hong Kong
promotor: prof.dr. S.J. Koopman
Contents

1 Introduction ........................................... 1
   1.1 Filtering time series .............................. 1
   1.2 State space models in econometrics .............. 2
   1.3 Non-linear state space models ...................... 4
   1.4 Model formulation ................................ 4
   1.5 Recursive Bayesian estimation ..................... 7
   1.6 Parameter estimation ............................... 9
   1.7 Thesis outline ................................ 11

2 Deterministic Filtering Algorithms ................. 13
   2.1 Introduction ..................................... 13
   2.2 Linear state space models ......................... 13
      2.2.1 The Kalman filter ............................. 14
      2.2.2 Smoothing and parameter estimation .......... 17
      2.2.3 Diffuse initialization ........................ 19
   2.3 The extended Kalman filter ......................... 21
      2.3.1 Deriving the extended filter ................. 21
      2.3.2 Smoothing and parameter estimation .......... 22
      2.3.3 Diffuse extended filter ...................... 23
   2.4 The unscented Kalman filter ....................... 23
      2.4.1 Unscented transformation ..................... 24
      2.4.2 Sigma point selection ....................... 26
      2.4.3 Unscented filtering summary, smoothing and parameter estimation 27
   2.5 State estimation error comparisons ............... 28
   2.6 Alternative methods ................................ 28

3 Monte Carlo Filtering Algorithms .................. 31
   3.1 Introduction ..................................... 31
   3.2 Importance sampling .............................. 32
3.2.1 Importance densities from linear Gaussian state space approximation ........................................... 33
3.2.2 Parameter estimation using importance sampling .......................................................... 34
3.2.3 Diagnostic testing for importance sampling ................................................................. 36
3.3 Particle filtering ................................................. 37
3.3.1 Generic particle filter .................................................. 37
3.3.2 Extended and unscented particle filters ................................................................. 38

4 Cycles with Smooth-Transition Parameters 41
4.1 Introduction .......................................................... 41
4.2 Trend-cycle decomposition model .......................................................... 44
  4.2.1 Fixed parameter specification .................................................. 44
  4.2.2 Time-varying parameter specification .................................................. 45
4.3 State space representation .................................................. 47
4.4 Empirical evidence from U.S. economic time series .................................................. 48
  4.4.1 Data .......................................................... 48
  4.4.2 Basic decompositions .................................................. 50
  4.4.3 Smooth transitions over time .................................................. 51
4.5 Discussion and conclusion .................................................. 61

5 Asymmetric Business Cycles 63
5.1 Introduction .......................................................... 63
5.2 Formulation of asymmetric stochastic cycle .......................................................... 65
  5.2.1 Asymmetric deterministic cycles .................................................. 66
  5.2.2 Asymmetric stochastic cycles .................................................. 68
5.3 Estimation and measurement .................................................. 69
  5.3.1 Trend-cycle decomposition model .................................................. 69
  5.3.2 State space form .................................................. 70
  5.3.3 Importance sampling .................................................. 71
  5.3.4 Importance weights .................................................. 72
5.4 Empirical evidence from U.S. economic time series .................................................. 73
  5.4.1 Data description .................................................. 73
  5.4.2 Parameter estimation for symmetric decomposition model .................................................. 73
  5.4.3 Parameter estimation for asymmetric decomposition model .............................................. 74
  5.4.4 Empirical evidence of asymmetric cycles .................................................. 76
5.5 Conclusion .......................................................... 78
6 Seasonality with Trend and Cycle Interactions

6.1 Introduction ........................................ 81
6.2 The unobserved components time series model .......... 83
6.3 Seasonal interacting components ........................ 86
6.3.1 A review of non-linear trend-seasonal models ......... 86
6.3.2 Trend and cycle interactions in the basic structural model 88
6.3.3 Seasonal interaction model in state space form ........ 89
6.3.4 Parameter estimation by the extended Kalman filter .... 90
6.4 Applications ................................... 91
6.4.1 U.K. visits abroad ................................ 91
6.4.2 U.S. unemployment .............................. 94
6.4.3 U.S. industrial and dwellings production ............. 97
6.5 Conclusion .................................. 100

7 Stochastic Volatility Models ............................. 103

7.1 Introduction ......................................... 103
7.2 Model specification and estimation ....................... 104
7.3 Likelihood estimation by importance sampling .......... 106
7.4 Obtaining importance densities ........................ 107
7.5 Simulation results .................................. 109
7.6 Importance sampling diagnostics ....................... 111
7.7 Stochastic volatility with t disturbances ............... 111
7.8 Conclusion .................................. 114

8 Conclusions ........................................... 119

8.1 General remarks ................................... 119
8.2 Methods ....................................... 119
8.3 Applications ................................... 120

Bibliography ........................................ 123

Samenvatting (Summary in Dutch) .................... 135

Acknowledgements .................................. 139
Chapter 1

Introduction

1.1 Filtering time series

A filter is a device which separates entities into their constituting components. The term originally described physical devices such as membranes that remove pollution from liquids or gases. In the twentieth century, electronic circuits were developed for the purpose of enhancing communication signals by separating transmission noise from the signal. Such electronic devices, and the mathematical models describing their characteristics were also termed filters. As the abstraction level of mathematical filtering models increased, they were adopted in a wide array of scientific disciplines. This thesis describes a selection of filtering procedures and applies them to economic time series. In common with physical filters, they will be used to separate elements from mixtures of components, although these components are more abstract entities. An important example is the separation of movements in economic activity into a slowly changing general direction (trend), cyclical variation of medium duration (business cycle), predictably repeating fluctuations throughout each year (season), and unpredictable movements without permanent impact (noise).

The filtering algorithms that form the subject of this thesis are generally applicable to time series models of a specific form. These models aim to represent a sequence of chronologically arranged numerical observations of some phenomenon as realisations of stochastic processes in discrete time which are indirectly observed, and often polluted by random disturbances. This class of models is referred to as a state space model, and enjoyed a surge in popularity in engineering and related disciplines following the publication of the Kalman filter algorithm (Kalman, 1960). Mathematical filters are used to reconstruct or estimate the indirectly observed stochastic processes from the observations by sequentially processing new observations when they become available. Additional algorithms are available to project the processes to periods in which ob-
servations are not (yet) available, yielding forecasts and interpolations. Due to the statistical nature of the filters, these projections are accompanied by measures of uncertainty which are generally not available from non-probabilistic interpolation and extrapolation algorithms.

Early applications of Kalman and related filtering techniques were found primarily in the problems of estimating and tracking the location of physical objects such as rockets and projectiles. Improving the clarity of imperfectly transmitted communication signals was another important utilisation. At an abstract level, filtering in state space models addresses the general problem of estimating indirectly observed dynamic systems. As such, it is not surprising that the state space approach has gained acceptance in a wide range of disciplines. Examples of applications outside communications and object tracking include computer image and speech recognition, navigation, machine learning, water level estimation, monitoring of seismic patterns, and a variety of economic and financial problems. Applications to economic time series form the focal points of the empirical sections of this thesis. However, the algorithms discussed in the methodological chapters are by no means limited, or even particularly specialised to such applications. Indeed, most of these methods found their origins outside economic and financial disciplines.

1.2 State space models in econometrics

As economic theory and practice is often concerned with latent or indirectly observed dynamic processes, it would appear that state space methodology is well suited to deal with economic estimation problems. However, over the past few decades econometric modelling of time series data has been dominated by the Box-Jenkins methodology. The authoritative tome developing this approach is the textbook *Time Series Analysis: Forecasting and Control* by Box and Jenkins (1970). In the Box-Jenkins approach, observations are first to be made stationary, that is, plausibly drawn from a process with time-invariant probabilistic properties. This is accomplished using various data transformations, such as taking logarithms and time differences. Subsequently autoregressive moving average (ARMA) models are fitted to the transformed data. The dynamic properties of the time series are usually assumed to be fully characterized by the serial correlation structure, and any stationary series can be modelled by a moving average polynomial of sufficient length which mimics the sampled serial correlations. A parsimonious representation of the time series is obtained by introducing an autoregressive polynomial, which combined with a low order moving average polynomial can account for a great variety of serial correlation patterns using a relatively small number of parameters.
The Box-Jenkins approach comprises the majority of the time series analysis component of econometric text books and university lectures. Further, wide availability of computer software, competitive forecasting performance compared with far more elaborate structural equations models, and a well developed mathematical theory have all contributed to the enduring popularity of this approach. The state space methodology however, offers a number of important advantages over ARMA type of models. The fact that state space methods do not require the data to be stationary eliminates a problem on which Box-Jenkins theory offers only minimal guidance. From the outset, state space methods are developed in a multivariate framework. They deal with missing and irregularly spaced observations in a straightforward way. Incorporating various desirable features such as interventions, breakpoints and regression effects are easy and natural. Finally, economic state space models are frequently built from components with direct interpretations. In contrast, ARMA models represent the aggregate dynamics of a multitude of stochastic components, and are often only employed as forecasting devices.

Early econometric applications of Kalman filtering ideas were concentrated on estimating regression models with time varying parameters. Articles developing maximum likelihood estimation in Gaussian state space models (Schweppe, 1965) and linking Kalman filtering to linear regression (Duncan and Horn, 1972) provided important bridges from the engineering disciplines to econometrics. During the late 1970s and early 1980s, more applications of Kalman filtering gradually appeared in the economic literature. Some notable contributions include articles in volume 2, issue 4 of *Annals of Economic and Social Measurement* (Berg, 1973), Bayesian filtering in Harrison and Stevens (1976), calculating the likelihood of ARMA models via the Kalman filter in Harvey and Phillips (1979), and the development of structural time series, or unobserved components models in a state space framework in Harvey and Todd (1983), Harvey (1984). Papers presented at the fifth World Congress of the Econometric Society by Harvey (1985) and Engle and Watson (1985) provide early overviews of Kalman filtering in econometrics. A broad historic context of recursive estimation in econometrics reaching back to the nineteenth century is provided by Pollock (2003). Financial applications, which typically utilise far more observations than other economic time series, were made possible by the rapid increase in computational resources since the late 1980s. Some examples of financial applications of state space models include estimation of capital asset pricing models (Wells, 1996), stochastic volatility (Ruiz, 1994), commodity prices (Schwartz, 1997), durations models (Bauwens and Veredas, 2004) and dynamic term structure models (Koopman et al., 2010).
1.3 Non-linear state space models

The focus of this thesis lies in non-linear models and methods. Although linear state space models have found numerous applications, the assumption of model linearity is an important limitation. Soon after Kalman’s publication in 1960, engineers have sought to overcome this restriction with various techniques. One of the earliest practical application of the Kalman filter involved navigation and control of the Apollo space capsule at NASA, which featured non-linear dynamics. As accounted by McGee and Schmidt (1985), this limitation was overcome by linearising the model, resulting in the extended Kalman filter, which remains to this day probably the most widely used non-linear filter.

In the 1960s and 70s, most of the research in non-linear filtering was concentrated on functional approximations to non-linear models. This led to the development of many variations on extended Kalman filters and other local approximations. The accuracy of these methods was further enhanced using mixtures of probability densities and iterative procedures.

The general filtering problem normally involves solving high dimensional integrals. Classical grid based numerical integration methods, using various quadrature formulas were naturally considered in the literature. However, it was widely recognised that these did not scale will beyond low dimensional problems. In most cases Monte Carlo or simulation methods have become the preferred approach. Although numerical integration using simulation techniques had been developed even before the introduction of Kalman filtering, application became only widespread after the personal computer revolution starting in the late 1970s. Much of the research on non-linear filtering since the late 1980s have focused on various simulation methods. Currently these appear to provide the most complete and satisfactory solutions to general non-linear filtering problems.

1.4 Model formulation

A concrete mathematical formulation of the non-linear state space models that form the subject of this thesis takes the form of the two equations

\[
\begin{align*}
\alpha_{t+1} &= T_t(\alpha_t) + \eta_t, \\
y_t &= Z_t(\alpha_t) + \epsilon_t,
\end{align*}
\]

for \( t = 1, \ldots, n \). The first equation is known as the state, transition, motion or process equation, or alternatively as the signal model or the dynamic model. It describes the evolution of the state vector \( \alpha_t \) as a first order Markov process, transformed from the
state in the previous time period by the deterministic vector valued function $T_t(\cdot)$, which may itself change (deterministically) over time. The state vector contains the variables required to predict the evolution of the system from one period to the next. Although few univariate series are “naturally” Markovian (that is, having distributions of the future that, conditioned on the present are independent of the past) they can often be reformulated as such in vector form by including a small amount of the history of the series in the form of lags as elements in the contemporaneous state vector.

The second equation is called the observation or measurement equation, and relates the vector of observed data $y_t$ to the unobserved state vector $\alpha_t$ through a (possibly time-varying) deterministic function $Z_t(\cdot)$. Thus, a state space model is a model for both the motion and the observation of a dynamic system. It is worth emphasizing that both the motion and the observation are vector processes, usually with different dimensions. (The dimensions may even vary in each time period $t$, although such generality is not encountered in the present work.) The versatility of the general state space formulation should be apparent. State space model are sometimes also called hidden Markov models. Although this terminology appears to be more descriptive, in the statistics and econometrics literature it is primarily associated with state space models with state variables which take on discrete values.

The disturbances $\eta_t$ are called the state, transition, or process noise and $\epsilon_t$ the observation or measurement noise. In most common specifications, and in this thesis, they are modelled as white noise, that is, zero mean, mutually and serially independent stochastic processes. Often, Gaussian or other specific distributions are assumed for the disturbance processes. The term $Z_t(\alpha_t)$ is sometimes referred to as the signal, so that an observation is the sum of signal and noise. Filtering results are also available for models with contemporaneous correlation between the observation and the measurement noise, and models with serial correlation in either noise process. In the latter case they are called coloured noise. Generally such models can be reformulated in a state space form with white noise by extending the state vector. Therefore they will not be dealt with explicitly in this thesis.

As usual in recursive model formulations, a complete model definition requires the specification of the initial conditions. These can be left implicit, as is often the case in ARMA modelling. The initial conditions are then derived from other assumptions, such as stationarity of the process. In state space model formulations they are usually explicit and given in the form of a distribution of the initial state vector $\alpha_1$.

Apart from trivial differences in notation, many variant descriptions of state space models are found in the literature which differ slightly from (1.1). A few common variations are noted here.

The time subscript in the transition equation can be shifted relative to that in the
observation equation, such that (1.1a) becomes
\[ \alpha_t = T_t(\alpha_{t-1}) + \eta_t, \] (1.2)
while the transition equations remains as in (1.1b). In models without dependence between observation and transition disturbances, this merely amounts to a difference in notation, rather than a difference in the model. However, when the transition and observation disturbances are correlated, formulations with different offsets in the time index should be handled with more care. Such models can generally be restated in the form of (1.1) by placing observation disturbances in the state.

State space models with non-additive disturbances can be written as
\[ \alpha_{t+1} = T_t(\alpha_t, \eta_t), \] (1.3a)
\[ y_t = Z_t(\alpha_t, \epsilon_t). \] (1.3b)

Although this formulation appears more general than the form with additive disturbances, it can also generally be reformulated in the additive form by placing the non-additive disturbances in the state vector. The additive disturbance form is the most common case in applications, and it is the preferred formulation in this thesis.

In the engineering literature, extra deterministic input in the form of control variables \( x_t \) is often added to the observation equation. These are used by the modeller to influence or control the system, resulting in the observation equation
\[ y_t = Z_t(\alpha_t, x_t) + \epsilon_t. \] (1.4)
In economics, similar formulations can be used to explicitly distinguish variables such as interventions, break point and treatment effects, which are assumed not to follow dynamic processes. However, these can again be included in the state. A possible implementation issue is that deterministic state variables do not have associated disturbance terms. This requires some modifications of many basic filter equations to avert singularity in the variance matrix of the disturbances.

A formulation which we will use in a later chapter defines the state equation as
\[ \alpha_{t+1} = T_t(\alpha_t) + R_t(\xi_t), \] (1.5)
where \( R_t(\cdot) \) is another deterministic transformation. This formulation has the advantage that the disturbance vector and the state vector can have different dimensions. Thus, the disturbance term can be defined with a non-singular variance matrix, even in models in which some components of the state have no associated disturbance components. The small adjustments required in the filtering algorithms will be noted where appropriate.
Finally, rather than giving equations transforming the state vector, non-linear state space models can be specified in the form of conditional probability distributions. Assuming the state distributions have properly defined densities, a general state space models can be specified by the transition and observation density

\begin{align}
\label{eq:1.6a}
p(\alpha_{t+1}|\alpha_{1:t}, y_{1:t}) &= p(\alpha_{t+1}|\alpha_t), \\
\label{eq:1.6b}
p(y_t|\alpha_{1:t}, y_{1:t-1}) &= p(y_t|\alpha_t),
\end{align}

respectively. The notation \(x_{i:j}\) is used to denote a set of variables or observations between \(i\) and \(j\), that is \(\{x_i, \ldots, x_j\}\). In order to keep notation simple, the convention of using \(p(\cdot)\) to represent a probability density of its argument(s) is used. Thus, \(p(\cdot)\) can represent densities from different distributions in the same context. When this might lead to ambiguity, densities will be subscripted with the symbol of the random variable or vector it represents.

The equations \eqref{eq:1.6} convey the notion that the state is a Markov process, and that conditional on the current state, the distribution of the current observation and future state is independent of the past. In other words, the current state summarises all the relevant probabilistic information from the past. This is the crucial aspect of state space modelling that enables efficient estimation using recursive equations which do not require storing and processing the rest of the history of the state or the observations in each step.

It is usually more straightforward to specify models in the form of functions as in \eqref{eq:1.1} than directly in probability densities. If the disturbance terms of the functional specification are assumed to have densities \(p_\eta(\cdot), p_\epsilon(\cdot)\) the density formulation of the model corresponding to \eqref{eq:1.1} is

\begin{align}
\label{eq:1.7a}
p(\alpha_{t+1}|\alpha_t) &= p_\eta(\alpha_{t+1} - T_t(\alpha_t)), \\
\label{eq:1.7b}
p(y_t|\alpha_t) &= p_\epsilon(y_t - Z_t(\alpha_t)).
\end{align}

When the disturbance processes are modelled as Gaussian white noise \(\eta_t \sim \text{NID}(0, H_t), \epsilon_t \sim \text{NID}(0, G_t)\), the non-linear state space model can be represented by

\begin{align}
\alpha_{t+1}|\alpha_t &\sim \text{NID}(T_t(\alpha_t), H_t), \\
y_t|\alpha_t &\sim \text{NID}(Z_t(\alpha_t), G_t).
\end{align}

\section{1.5 Recursive Bayesian estimation}

The main problem to be solved in state space modelling is the estimation of the unobserved quantities in the model, which are the state vectors \(\alpha_1, \ldots, \alpha_n\) and model
parameters \( \theta \) upon which the observation and transition equations or densities may depend. Assuming for the moment that \( \theta \) and the initial state distributions are known, a complete recursive solution to the state estimation problem can be stated in the form of a prediction equation and a filtering or updating equation. These two recursive equations alternatively compute the densities of the current and the first future state, conditional on all available observations. This solution is known as Bayesian filtering or recursive Bayesian estimation, and is attributed to Ho and Lee (1964).

Using the shorthand notation \( Y_t = y_{1:t} \), the prediction equation is given by

\[
\begin{align*}
p(\alpha_{t+1}|Y_t) &= \int p(\alpha_{t+1}, \alpha_t|Y_t) \, d\alpha_t \\
&= \int p(\alpha_{t+1}|\alpha_t)p(\alpha_t|Y_t) \, d\alpha_t
\end{align*}
\] (1.9)

and takes as input the transition density (1.6a) and the filtering density (1.10). The equation follows directly from applying Bayes’ rule and observing that \( \alpha_t \) contains all the relevant information from the observations up to the current time to predict \( \alpha_{t+1} \).

The filtering equation produces the density

\[
p(\alpha_t|Y_t) = \frac{p(\alpha_t, y_t|Y_{t-1})}{p(y_t|Y_{t-1})} = \frac{p(\alpha_t|Y_{t-1})p(y_t|\alpha_t|Y_{t-1})}{\int p(y_t|\alpha_t)p(\alpha_t|Y_{t-1}) \, d\alpha_t}
\] (1.10)

from the observation density (1.6b) and the prediction density (1.9) of the previous period by application of Bayes’ rule. These recursive equations are initialised with a given density of the initial state \( p(\alpha_1|Y_0) = p(\alpha_1) \).

If a state estimator is conditioned on observations which include values from the future relative to the period of the state to be estimated, it is called a smoother. Although different sets of observations can be considered, in economic applications (and throughout this thesis), smoothing usually refers to state estimation based on the entire data set \( Y_n \). A recursive equation for the smoothing density of the state at time \( t \) based on the entire data set is given by

\[
p(\alpha_t|Y_n) = \int p(\alpha_t, \alpha_{t+1}|Y_n) \, d\alpha_{t+1} = \int p(\alpha_{t+1}|Y_n)p(\alpha_t|\alpha_{t+1}, Y_t) \, d\alpha_{t+1}
\] (1.11)

which follows again from repeated application of Bayes’ formula and the Markov property of the state. Starting from the final density \( p(\alpha_n|Y_n) \) of the filtering recursion (1.10),
1.6 Parameter estimation

The general state space model (1.1) or (1.7) is in practical applications usually specified up to some fixed parameters. The most common examples are the variances and covariances of the disturbance terms, but the transition and observation functions $T_t(\cdot)$ and $Z_t(\cdot)$ can also contain unknown parameters. In economic applications of state space models, estimation of such parameters often receives as much attention as estimation of the state process. This contrasts to engineering and physical applications, which usually put an emphasis on (real-time) state estimation.

In linear state space modelling, the dominant approach to estimation of unknown fixed parameters is the maximum likelihood method. In principle the parameters could be estimated using the recursive Bayesian technique described in the previous section by including them in the state vector with trivial transition equations and disturbances. There is a good precedence for such an approach, since some of the earliest applications of Kalman filtering to economic time series involved estimating regression models with time varying coefficients models, where fixed linear regression coefficients are put in a state vector. Economic theory rarely provides a useful guide to distinguish between fixed parameters and dynamic latent processes. In linear state space modelling, one could adopt a rather pragmatical rule of assuming variables which would render the model non-linear to be fixed parameters. When the state space framework to be adopted is
explicitly non-linear, such a distinction between state variables and parameters seems contrived. In the engineering literature, estimating parameters as part of the state vector is more common and often termed on-line parameter learning or adaptive parameter estimation. However, there are limitations both theoretical and practical to including all unknown variables in a state vector. In formulating economic theory, some sense of model stability is usually desirable, and fixed parameters provide a natural means to accomplish this. More practically, if fixed parameters are included in the state without accompanying disturbance terms (or equivalently, imposing zero variance on the disturbance terms), the problem shifts from estimating unknown parameters to estimating values of unknown initial state components. If they are included and modelled as time varying processes, they gain new (variance) parameters which in their turn need to be estimated. Further, when state variables are assumed to be fixed, the estimates can be quite sensitive to the imperfections of non-linear estimation algorithms. The basic approach to develop practical non-linear models in this thesis is to furnish existing linear models with modest non-linear extensions and modifications. Hence maximum likelihood methods as widely used in linear state space modelling are also adopted as the primary parameter estimation technique.

The likelihood function in state space models can be efficiently calculated from the filtering recursions by breaking the joint density of the observations into densities of current observations conditioned on the past. This is called the prediction error decomposition (Schweppe, 1965), and presents the log-likelihood function in the form

$$\log L(\theta) = \sum_{t=1}^{n} \log p(y_t|Y_{t-1}),$$

where parameters $\theta$ are implicit in the densities $p(\cdot)$. The terms

$$p(y_t|Y_{t-1}) = \int p(y_t|\alpha_t)p(\alpha_t|Y_{t-1}) \, d\alpha_t$$

in the summation appear in the denominator of (1.10) when calculating the filter for a given set of parameters, so that the likelihood is easily obtained as a by-product of the filter calculations.

The Bayesian filtering and parameter estimation solutions are conceptual in the sense that they don’t provide practically implementable algorithms for most applications. In the general case, the solution requires the storage of and the repeated integration over arbitrary probability densities of potentially high dimension. Models with linear transition and disturbance equations and Gaussian disturbances are an important special case which can be solved explicitly and completely. The main subjects of this thesis are models of economic time series which fit in the non-linear state space form, and practical approximating algorithms to estimate such models.
1.7 Thesis outline

The remainder of this thesis is organised as follows. In chapter 2, deterministic state space estimation methods are described. The basic linear Kalman filter is derived in a Gaussian model. The extended and unscented filter and smoother are developed to estimate non-linear models.

Chapter 3 describes simulation based estimation methods. Details are provided for global and sequential importance sampling techniques. A number of alternatives are briefly described.

The first three chapters constitute the methodological basis of this thesis. The next four chapters contain selected applications of non-linear filtering. These are substantially based on the papers Koopman and Lee (2004), Koopman and Lee (2005), Koopman, Lee, and Wong (2006) and Koopman and Lee (2009).

Chapter 4 and 5 develop extensions of the cycle specification in standard unobserved components models to account for dynamic parameters. Specifically, smooth transition functions are applied to the cycle variance and period parameters, and the length of the period is made phase dependent. The models are applied to various U.S. macro economic time series.

Chapter 6 introduces UC models with interactions between components. The seasonal component is allowed to vary depending on the trend, and in different phases of the business cycle. Illustrations are provided for U.K. travel data and U.S. unemployment and production series.

In chapter 7 two importance sampling based estimations methods for stochastic volatility models are investigated. The parameter estimation performance of two approximation models are compared is a simulation study.

Chapter 4 and 6 use the (diffuse) linear and extended Kalman filter techniques discussed in chapter 3. The non-linear models of chapter 5 and 7 are estimated using simulation methods elaborated in Chapter 3. Chapter 8 contains a general summary and conclusion of this thesis.
CHAPTER 1. INTRODUCTION
Chapter 2

Deterministic Filtering Algorithms

2.1 Introduction

This chapter describes widely used algorithms employed in estimating the non-linear state space model introduced in section 1.4. The methods in this chapter are deterministic, in the sense that they do not rely on sequences of (pseudo-)random numbers. The classic linear and extended Kalman filter are reviewed with great detail, as well as the unscented Kalman filter which has become popular in recent engineering publications. Finally, brief descriptions are provided of some prominent alternative methods.

2.2 Linear state space models

The linear state space model is an important special case of the general state space model described in the previous chapter. When it is combined with Gaussian distribution assumptions for the disturbances and the initial state, the Kalman filter provides an exact and complete solution to the Bayesian filtering problem which is very economical in computer calculation and storage requirements. The filter prediction and update steps require a few basic matrix calculations, and only the conditional means and variances of the filtering and prediction density need to be stored in each step of the iteration.

The linear state space model is a special form of the general model (1.1) in which the observation and transition equations are both linear. The equations are given by

\begin{align*}
\alpha_{t+1} &= T_t \alpha_t + \eta_t, \tag{2.1a} \\
y_t &= Z_t \alpha_t + \epsilon_t, \tag{2.1b}
\end{align*}

in which the transition matrices \( T_t \) and the observation matrices \( Z_t \) may depend on unknown parameters. The disturbances are assumed to be Gaussian white noise \( \epsilon_t \sim \)
It is noteworthy that the linear state space model encompasses two models which are widely used in economics. The state vector follows a first order vector auto-regressive model, which in its turn generalises scalar ARMA models. All univariate ARMA models can be written as a linear state space model, and conversely many models originally formulated in state space form have a known ARMA representation. A detailed overview is provided by Harvey (1989). The observation equation (2.1b) resembles a linear regression model, with the distinction that the coefficients are stochastic dynamic processes rather than fixed parameters. Many more parallels can be drawn between linear regression and linear state space models, and between least squares estimates and the Kalman filter; some of these will be noted in the passing in this section.

Although far more restrictive than the general non-linear model, the linear model has been widely applied to many economic and time series problems. Like the linear regression model, the linearity assumptions are not quite as confining as they might appear at first. The observations $y_t$ may be non-linearly transformed before the model is applied, and the elements of $Z_t$ can be subject to non-linear transformations from data, analogous to exogenous variables in linear regression models.

2.2.1 The Kalman filter

The Kalman recursions have been derived in different ways starting from different sets of assumptions in many textbooks and articles. Similar to the least squares algorithm in regression models, it can be derived from probabilistic or geometric principles. Assuming Gaussianity for the initial state and for the disturbance terms, a probabilistic derivation can proceed from the general Bayesian recursion from section 1.5 and writing out the Gaussian densities. This approach is taken in Ho and Lee (1964), Kitagawa and Gersch (1996). In a geometrically oriented derivation, the stochastic vectors in the state space are modelled as vectors in a Hilbert space. Orthogonality in this space is defined as uncorrelatedness between vectors. Using the Projection Theorem, the filter for the state is derived as the projection of the state on the space spanned by the past observations. Proofs following this approach can be found in Kalman (1960), Brockwell and Davis (1991) and do not require distributional assumptions, other than that the random vectors have finite second moments. Assuming linearity from the start, the filter can also be directly derived from standard results from multivariate Gaussian distribution theory, as in Anderson and Moore (1979), Harvey (1989) or Durbin and Koopman (2001). We follow this approach here.

If $x$ and $y$ are jointly Gaussian with $x \sim N(\mu_x, \Sigma_{xx})$, $y \sim N(\mu_y, \Sigma_{yy})$ and $\text{Cov}(x, y) = \Sigma_{xy}$, the conditional distribution of $x|y$ is Gaussian with expectation and variance given
by

\[ E(x|y) = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y), \] (2.2a)
\[ \text{Var}(x|y) = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}, \] (2.2b)

assuming that \( \Sigma_{yy} \) is non-singular. This result can be found in basic statistics and econometrics textbooks, see for instance Greene (2003), Theorem B.7.

To derive a recursion for \( E(\alpha_t|Y_t) \) and \( \text{Var}(\alpha_t|Y_t) \), we first define \( v_t = y_t - E(y_t|Y_{t-1}) \). Using the law of iterated expectations, it follows immediately that \( E(v_t) = 0 \). We assume that \( \alpha_t|Y_{t-1} \) and \( v_t \) are jointly Gaussian. This is easily shown to hold if the state space model is linear as in (2.1) with Gaussian disturbance terms, and the initial state distribution \( \alpha_1|Y_0 = \alpha_1 \) is also Gaussian. It is clear that conditioning the state on \( \{v_t, Y_{t-1}\} \) is equivalent to conditioning the state on \( Y_t \), from which it follows from (2.2) that

\[ E(\alpha_t|Y_t) = E(\alpha_t|Y_{t-1}, v_t) = E(\alpha_t|Y_{t-1}) + K_t v_t \] (2.3)

and

\[ \text{Var}(\alpha_t|Y_t) = \text{Var}(\alpha_t|Y_{t-1}) - K_t \times \text{Var}(v_t) \times K_t', \] (2.4)

where

\[ K_t = \text{Cov}(\alpha_t, v_t|Y_{t-1}) \times [\text{Var}(v_t)]^{-1}. \] (2.5)

Define the variables

\[ a_t = E(\alpha_t|Y_{t-1}), \quad P_t = \text{Var}(\alpha_t|Y_{t-1}). \] (2.6)

In the linear model, we have

\[ v_t = y_t - E(Z_t\alpha_t + \epsilon_t|Y_{t-1}) = y_t - Z_t a_t \]
\[ \text{Var}(v_t) = \text{Var}(Z_t\alpha_t - Z_t a_t + \epsilon_t|Y_{t-1}) = Z_t P_t Z_t' + G_t, \] (2.7)
\[ \text{Cov}(\alpha_t, v_t|Y_{t-1}) = \text{E}(\alpha_t - a_t)(Z_t\alpha_t - Z_t a_t + \epsilon_t)'|Y_{t-1}) = P_t Z_t', \]

Substituting these into (2.3–2.5) yield formulas for the filtered expectation and variance.
The prediction step follows directly from the assumed transition equation:

\[ a_{t+1} = E(\alpha_{t+1}|Y_t) = E(T_t \alpha_t + \eta_t|Y_t) = T_t a_t|t, \]
\[ P_{t+1} = \text{Var}(\alpha_{t+1}|Y_t) = \text{Var}(T_t \alpha_t + \eta_t|Y_t) = T_t P_t|t T_t' + H_t, \]

where

\[ a_{t|t} = E(\alpha_t|Y_t), \quad P_{t|t} = \text{Var}(\alpha_t|Y_t), \]

were calculated in the filter step. Collecting these together, we obtain the Kalman recursions

\[

t_v = y_t - Z_t a_t, \quad F_t = Z_t P_t Z_t' + G_t, \\
K_t = P_t Z_t' F_t^{-1}, \quad P_{t|t} = P_t - K_t F_t K_t', \\
a_{t+1} = T_t a_t|t, \quad P_{t+1} = T_t P_t|t T_t' + H_t. 
\]

The recursions are initialised with \( a_1, P_1 \), which are defined as the unconditional mean and variance respectively of \( \alpha_1 \), and calculate the filtered state at time \( t \) and the predicted state of time \( t + 1 \) each time a new observation \( y_t \) arrives.

The variables \( v_t \) are one-step prediction errors or innovations, with associated variance matrices \( F_t \). In the filter equations above, they are primarily defined to facilitate the calculations. In a well fitting model, the standardised innovations sequence are white noise with unit variance. Hence, diagnostic tests of the assumed model are often performed on the innovations.

The variable \( K_t \) is known as the Kalman gain matrix, and has the interpretation of a weighting matrix used to update knowledge of the state by weighting new evidence (observations) against the assumed model. This intuition is especially clear in models with univariate observations and states. In that case, the scalar gain matrix decreases when the observation disturbance variances \( G_t \) are large relative to the state disturbance variances \( H_t \). Thus a new innovation from a very uncertain observation has only a small influence on the state estimate.

It is worth noting that in the equations (2.10) only \( v_t, a_{t|t} \) and \( a_{t+1} \) depend on \( y_t \). The Kalman gain and the variances do not depend on observations, but only on the assumed model and hence can be completely pre-calculated before any data is available. This property does not generalise to the non-linear algorithms, however.

If the disturbance and initial state have finite second moments but no further distributional assumptions are made, the Kalman filter provides the minimum linear mean
square error (MLMSE) estimates of the state. These are MMSE within the restrictive class of estimators which are calculated as linear transformations of the observations. Under Gaussianity assumptions, the conditional expectations of the filter are general MMSE estimates, as noted in section 1.6. The derivation under Gaussianity has the obvious disadvantage of providing little information about its performance when the Gaussianity assumption does not hold. However, in the latter case it is unclear how to produce a general MMSE estimate and to what degree it is superior to the MLMSE estimator. In his original derivation of the filter, Kalman noted that its interpretation as a general MMSE estimate for Gaussian processes or a MLMSE estimate for processes with less restrictive distributional assumptions can be considered to be largely a matter of taste.

### 2.2.2 Smoothing and parameter estimation

As defined in section 1.5, a smoother is an estimator of the state which incorporates future observations. In the state space literature, distinction is sometimes made between fixed point, fixed lag, and fixed interval smoothing. In fixed point smoothing, estimates of the states at some given points in time are continuously updated when more observations become available. Official economic statistics which are often revised several times after initial publication, are a natural application. In fixed lag smoothing, estimates on the current state are provided after some fixed number of future observations are available. This is mainly relevant when (near) real-time estimates are desired, such as in navigation applications, and the complete set of observations is very large. A fixed number of extra observations are used in fixed lag smoothing to improve estimation accuracy, at the cost of some time delay before having estimates available. In fixed interval smoothing, states are estimated conditioned on observations from a given time interval. In the most common case the interval is the entire period for which observations are available. This is the most relevant smoothing concept in economics, as research is typically conducted using data sets of limited and fixed sizes which are not frequently updated. It is also the only smoothing concept discussed in this thesis.

In contrast to filtering and prediction, a number of quite different algorithms are available to calculate the solution to the smoothing problem. All of the smoothing methods involve at least an additional backward recursive pass through the data, after the forward pass of the Kalman filter. The classical fixed interval smoother derived by Rauch et al. (1965) and later by Ansley and Kohn (1982) directly modifies the previously calculated filtered states and variances. de Jong (1988) provides a more efficient algorithm which avoids the potentially costly inversion of the state prediction variance matrix. Another type of smoothers is derived by combining the forward
Kalman filter output with the output of a Kalman filter run backwards using an inverse model (Mayne, 1966; Merkus et al., 1993). In the linear model, the results of these smoothing algorithms are identical (although in computer representations they may differ due to accumulated rounding errors).

A brief derivation of the classical interval smoother for the state is given here, which will be used later in the non-linear sections, followed by a description of the smoother of de Jong (1988). The smoothed state \( \hat{\alpha}_t \) is the expectation of the state at time \( t \) conditional on all observations, and can be obtained by marginalising the joint conditional distribution of \( \alpha_t, \alpha_{t+1} \) given \( Y_n \),

\[
\hat{\alpha}_t = \mathbb{E}(\alpha_t|Y_n),
\]

\[
= \int \mathbb{E}(\alpha_t|\alpha_{t+1}, Y_n)p(\alpha_{t+1}|Y_n) \, d\alpha_{t+1},
\]

\[
= \int \mathbb{E}(\alpha_t|\alpha_{t+1}, Y_t)p(\alpha_{t+1}|Y_n) \, d\alpha_{t+1},
\]

where the last equation follows from the assumed Markov structure of the state space model. Now consider the joint conditional distribution of \( \alpha_{t|t} \) and \( \alpha_{t+1|t} \) which have respective means \( a_{t|t} \), \( a_{t+1|t} \), variances \( P_{t|t} \), \( P_{t+1|t} \) and covariance \( M_t \). From (2.2) we derive

\[
\mathbb{E}(\alpha_t|\alpha_{t+1}, Y_t) = a_{t|t} + M_t P_{t+1|t}^{-1}(\alpha_{t+1|t} - a_{t+1}),
\]

where we write \( \alpha_{t+1|t} \) is the conditional random vector \( \alpha_{t+1} \) given \( Y_t \). Substituting into (2.11) yields

\[
\hat{\alpha}_t = \int (a_{t|t} + M_t P_{t+1|t}^{-1}(\alpha_{t+1|t} - a_{t+1}))p(\alpha_{t+1|t}|Y_n) \, d\alpha_{t+1} \quad \text{(2.13)}
\]

where in the linear Gaussian model \( a_{t|t} \) and \( a_{t+1} \) can be obtained from the Kalman filter and

\[
M_t = \text{Cov}(\alpha_t, \alpha_{t+1}|Y_t) = \mathbb{E}(\left(\alpha_t - a_{t|t}\right)\left(\alpha_{t+1} - a_{t+1}\right)^T|Y_t),
\]

\[
= P_{t|t} T_t^T,
\]

where the covariance in the first equation is between the state variables conditioned on \( Y_t \). The backwards recursion is initialised by the last filtered state \( \hat{\alpha}_n = a_{n|n} \). Similarly, a recursion for the smoothed variance can be derived in the form of

\[
V_t = \text{Var}(\alpha_t|Y_n) = P_{t|t} - M_t P_{t+1}^{-1}(P_{t+1} - V_{t+1}) P_{t+1}^{-1} M_t^T,
\]

where

\[
W_t = \int a_{t+1|t}^T p(\alpha_{t+1|t}|Y_n) \, d\alpha_{t+1},
\]

\[
\hat{\alpha}_t = \frac{a_{t+1|t}}{W_t}.
\]
The basic state smoothing algorithm of de Jong (1988) is given without proof here. These take the form of the backwards recursions

\[ L_t = T_t - K_t Z_t, \quad r_{t-1} = Z_t' F_t^{-1} v_t + L_t' r_t, \quad N_{t-1} = Z_t' F_t^{-1} Z_t + L_t' N_t L_t, \]  
\[(2.16)\]

which are are initialised with \( r_n = 0, N_n = 0 \). Afterwards the smoothed state and variance are calculated as

\[ \hat{\alpha}_t = a_t + P_t r_{t-1}, \quad V_t = P_t - P_t N_{t-1} P_t, \]  
\[(2.17)\]

respectively for \( t = 1, \ldots, n \).

The Gaussian likelihood function of the linear state space model is given by

\[ \log L(\theta) = -\frac{n}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{n} \log |F_t| - \frac{1}{2} \sum_{t=1}^{n} v_t' F_t^{-1} v_t. \]  
\[(2.18)\]

This is calculated from the standardised innovations of the Kalman filter, which depend on the parameter vector \( \theta \). It is readily derived from the general prediction error decomposition in section 1.5, the substitution the Gaussian density and the expressions for the conditional expectations. Operationally, linear state space estimation is started by using a numerical optimization algorithm to find a parameter vector \( \hat{\theta} \) that maximises (2.18). In many state space methods, the exact score can be calculated, so that quasi-Newton algorithms are very suitable optimisation methods, see Koopman and Shephard (1992). An alternative parameter estimation method which is often faster than numerical likelihood optimization is the expectation maximization algorithm of Dempster et al. (1977). Its use in the state space context have been developed by Shumway and Stoffer (1982) and Watson and Engle (1983). Expectation maximization can also be used to produce initial values to use in numerical likelihood maximization. After parameter estimates are found, statistical tests are performed on the innovations from the filter evaluated with the estimated parameters to detect potential model miss-specifications. Finally, the state is estimated using the smoothing recursions, assuming that the estimated parameter vector is correct.

### 2.2.3 Diffuse initialization

Although state space modelling does not require the observation or state process to be stationary, this freedom carries a price. When the variables in the state are stationary, the Kalman filter can be initialised by the mean and covariance matrix of the unconditional distribution of the state. For non-stationary variables, the initial conditions are not obvious. When theory or other external considerations do not suggest an initial distribution, it can be modelled with a flat or non-informative distribution. The
simplest way to implement this computationally is to assume a distribution with some arbitrary mean and a very large variances, such as \( \alpha_1 \sim N(0, kI) \), and start the filter calculations with a very large value for \( k \).

More precise methods have been devised to deal with this problem, which is known in the literature as \textit{diffuse initialization}, see de Jong (1991), Ansley and Kohn (1985). Here, we describe the initialisation method by Koopman and Durbin (2003), which builds on Koopman (1997) and Ansley and Kohn (1985).

When the state vector contains diffuse variables, the mean square error matrix \( P_t \) can be decomposed into a part associated with diffuse state elements \( P_{\infty,t} \), and a part where the state has a proper distribution \( P_{s,t} \), that is

\[
P_t = kP_{\infty,t} + P_{s,t}, \quad k \to \infty.
\]

(2.19)

For the diffuse initial state elements, the corresponding entries on the diagonal matrix \( P_{\infty,t} \) are set to positive values, while the remainder of the matrix contains zeros. Koopman and Durbin show that for models with diffuse state elements the standard Kalman filter can be split into two parts by expanding the inverse of \( F_t = kF_{\infty,t} + F_{s,t} + O(k^{-2}) \) in \( k^{-1} \), where \( F_{\infty,t} \) is assumed to be invertible if it is non-zero. In the first \( d \) iterations of the filter, the state contains diffuse elements, which is indicated by a non-zero \( P_{\infty,t} \). Separate update equations are maintained for the parts associated with \( P_{\infty,t} \) and with \( P_{s,t} \). Generally \( P_{\infty,t} \) becomes zero after some iterations, after which the standard Kalman filter can be used. Letting \( k \to \infty \), the diffuse filter equations for the initial iterations are given by

\[
\begin{align*}
v_t &= y_t - Z_t a_t, \\
F_{\infty,t} &= Z_t P_{\infty,t} Z_t', \\
K_{\infty,t} &= P_{\infty,t} Z_t' F_{\infty,t}^{-1}, \\
P_{\infty,t} &= P_{\infty,t} - K_{\infty,t} F_{\infty,t} K_{\infty,t}', \\
a_{t|t} &= a_t + K_{\infty,t} v_t \\
a_{t+1} &= T_t a_{t|t},
\end{align*}
\]

(2.20)

when \( F_{\infty,t} > 0 \). In case \( F_{\infty,t} \) is zero \( K_{\infty,t} \) does not exist, and the equations for \( K_{s,t}, P_{\infty,t}, P_{s,t} \) and \( a_{t|t} \) are given by

\[
\begin{align*}
K_{s,t} &= P_{s,t} Z_t' F_{s,t}^{-1}, \\
P_{\infty,t} &= P_{\infty,t}, \\
P_{s,t} &= P_{s,t} - K_{s,t} Z_t P_{s,t}', \\
a_{t|t} &= a_t + K_{s,t} v_t, \\
a_{t+1} &= T_t a_{t|t}.
\end{align*}
\]

(2.21)

The diffuse likelihood based on this filter is given by

\[
\log L(\theta) = -\frac{n}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{d} w_t - \frac{1}{2} \sum_{t=d}^{n} \log |F_t| - \frac{1}{2} \sum_{t=d}^{n} v_t' F_t^{-1} v_t
\]

(2.22)
2.3. THE EXTENDED KALMAN FILTER

2.3.1 Deriving the extended filter

The extended Kalman filter was developed by Stanley Schmidt at the NASA Ames Research Center for use in a non-linear navigation problem. After experimenting with a number of linearisation schemes, the most satisfactory results were obtained from linearising the transitions and observation equations around the most current estimate of the state. The operational approximation is the first order Taylor expansion of (1.1) given by

$$\alpha_{t+1} \approx T_t(a_{t|t}) + \dot{T}_t \cdot (\alpha_t - a_{t|t}) + \eta_t,$$

$$y_t \approx Z_t(a_t) + \dot{Z}_t \cdot (\alpha_t - a_t) + \epsilon_t,$$

where

$$\dot{Z}_t = \left. \frac{\partial Z(x)}{\partial x'} \right|_{x=a_t}, \quad \dot{T}_t = \left. \frac{\partial T(x)}{\partial x'} \right|_{x=a_{t|t}},$$

as the predicted and filtered states $a_t$ and $a_{t|t}$ respectively, are the most recent state estimates available when the linearisations are required to run the linear filter equations.

Since this approximation is linear in $\alpha_t$, the basic Kalman filter of (2.10) can be applied with some small modifications. If the linear state space model (2.1) is extended with non-random vectors $c_t, d_t$ as in

$$\alpha_{t+1} = d_t + T_t a_t + \eta_t,$$

$$y_t = c_t + Z_t a_t + \epsilon_t,$$

the equations for $v_t$ and $a_{t+1}$ in the Kalman filter (2.10) are still valid, when the equations for $v_t$ and $a_{t+1}$ are replaced as follows:

$$v_t = y_t - Z_t a_t - c_t,$$

$$a_{t+1} = T_t a_{t|t} + d_t.$$

This is easily seen since the deterministic terms only affect the means, but not the variances. Model (2.23) is a linear state space model of the form (2.25) with

$$d_t = T_t(a_{t|t}) - \dot{T}_t a_{t|t},$$

$$c_t = Z_t(a_t) - \dot{Z}_t a_t.$$
Substituting these in the modified linear Kalman filter yields

\[ v_t = y_t - Z_t(a_t), \quad F_t = \dot{Z}_t P_t \dot{Z}_t' + G_t, \]

\[ K_t = P_t \dot{Z}_t' F_t^{-1}, \]

\[ a_{t|t} = a_t + K_t v_t, \quad P_{t|t} = P_t - K_t F_t K_t', \]

\[ a_{t+1} = T_t(a_{t|t}), \quad P_{t+1} = \dot{T}_t P_{t|t} \dot{T}_t' + H_t, \]

which comprise the extended Kalman filter for the non-linear state space model (1.1).

The extended filter is widely used in location tracking and navigation applications. The method is a straightforward modification to the basic Kalman filter, with little additional computational or theoretical burden. However, it has a number of known drawbacks. In practice, when model non-linearities are severe, the filter can become unstable and diverge from the true state position after some iterations. From a functional perspective, a first order approximation can be quite crude. In probabilistic terms, the estimates are clearly biased, since \( E(f(X)) \neq f(E(X)) \) in general. Nonetheless, when the non-linearities in the model are modest, it is a simple and effective means to estimate the state. It emphasizes the value of the state (in physical applications often the location of some object) rather than its probability distribution.

### 2.3.2 Smoothing and parameter estimation

Extended smoothing algorithms can be developed in a similar manner as the extended filter. The classical fixed interval smoother of (2.13) is modified to yield

\[ \hat{\alpha}_{t+1} = a_{t|t} + P_{t|t} \hat{T}_t P_{t+1|t}^{-1} (\hat{\alpha}_{t+1} - a_{t+1}), \]

(2.30)

where all the quantities on the right hand side have been obtained from the forward pass of the extended filter.

Alternatively, applying the linear smoothing equations (2.16) to the approximating model, yields the backwards recursion

\[ L_t = \dot{T}_t - \dot{T}_t K_t \dot{Z}_t, \quad r_{t-1} = \dot{Z}_t F_t^{-1} v_t + L_t' r_t, \quad N_{t-1} = \dot{Z}_t' F_t^{-1} \dot{Z}_t + L_t' N_t L_t, \]

(2.31)

where the smoothed state estimate \( \hat{\alpha}_t \) and variance \( V_t \) are as in (2.17).

The likelihood approximation of the non-linear model can be obtained by using the same prediction error decomposition of (2.18) with the innovations and variances from the extended filter. Parameter estimates based on maximising this likelihood function can be interpreted as maximum likelihood estimates of the linearised model, or as approximate estimates of the non-linear model. However, general results on the approximation performance appear to be unavailable.
2.3.3 Diffuse extended filter

Since the extended filter can be regarded as a linear filter applied to approximating model (2.23), we can apply essentially the same method for a diffuse initialization. This yields the diffuse extended Kalman filter for $t = 1, \ldots, d$, given by

\begin{align*}
v_t &= y_t - Z_t(a_t), \\
F_{\infty,t} &= \dot{Z}_t P_{\infty,t} \dot{Z}_t', \\
K_{\infty,t} &= P_{\infty,t} \dot{Z}_t F_{\infty,t}^{-1}, \\
P_{\infty,t|t} &= T_t P_{\infty,t|t} T_t', \\
a_{t+1} &= a_t + K_{\infty,t} v_t, \\
F_{s,t} &= \dot{Z}_t P_{s,t} \dot{Z}_t' + \sigma^2, \\
K_{s,t} &= (P_{s,t} \dot{Z}_t' - K_{\infty,t} F_{s,t}) F_{\infty,t}^{-1}, \\
P_{s,t|t} &= P_{s,t} - K_{s,t} \dot{Z}_t P_{s,t} - K_{s,t} \dot{Z}_t P_{\infty,t|t}, \\
\sigma^2 &= 2, \\
T_t &= T_t(a_{t|t}),
\end{align*}

(2.32)

for $F_{\infty,t} > 0$, and

\begin{align*}
K_{s,t} &= P_{s,t} \dot{Z}_t' F_{s,t}^{-1}, \\
P_{s,d|t} &= P_{s,d|t}, \\
P_{s,t|t} &= P_{s,t} - K_{s,t} \dot{Z}_t P_{s,t}'
\end{align*}

(2.33)

for $F_{\infty,t} = 0$. When $P_{\infty,t}$ becomes zero after $d$ iterations, the standard extended filter of (2.29) applies with $P_d = P_{s,d}$. The approximate diffuse extended likelihood is given by (2.22) where $v_t, F_{\infty,t}, F_{s,t}$ are obtained from the diffuse extended filter.

2.4 The unscented Kalman filter

The extended Kalman filter has been the workhorse algorithm for non-linear state estimation in engineering application for decades. The more recently developed unscented Kalman filter (Julier and Uhlmann, 1997) has been shown to improve upon its performance with comparably low calculation complexity. This algorithm is based on a fast approximation of high order Taylor expansions of the mean and variance of non-linearly transformed random variables. The extended filter is essentially a first order approximation to model non-linearity. The unscented filter has two distinct advantages: it carries out the approximation to higher order terms and it does not require the calculation of any derivatives. Further, expanding the orders is relatively easy. In this section, we provide the basic unscented filter, leaving details of the approximation algorithm to the following subsections.

In the derivation of the standard Kalman filter in section 2.2.1, the first equations (2.3), (2.4), (2.5) do not depend on assumptions of linearity of the state or observation equations, but only on joint Gaussianity of the conditional state and observation densities. Consequently, if we assume Gaussianity of these densities in each step (rather
CHAPTER 2. DETERMINISTIC FILTERING ALGORITHMS

than assuming Gaussianity of all the disturbance terms), we still have

\begin{align*}
    a_{t|t} &= \mathbb{E}(\alpha_t|Y_t) = a_t + K_t v_t, \\
    P_{t|t} &= \text{Var}(\alpha_t|Y_t) = P_t - K_t \times \text{Var}(v_t|Y_{t-1}) \times K_t',
\end{align*}

(2.34)
as before, but with

\begin{align*}
    v_t &= y_t - \mathbb{E}(Z_t(\alpha_t)|Y_{t-1}), \\
    K_t &= \text{Cov}(\alpha_t, v_t|Y_{t-1}) \times [\text{Var}(v_t)]^{-1} \\
         &= \mathbb{E}((\alpha_t - a_t)(Z_t(\alpha_t) - \mathbb{E}[Z_t(\alpha_t)]|Y_{t-1}) \times [\text{Var}(Z_t(\alpha_t|Y_{t-1})) + G_t]^{-1},
\end{align*}

(2.35)
and predictions

\begin{align*}
    a_{t+1} &= \mathbb{E}(\alpha_{t+1}|Y_t) \\
    &= \mathbb{E}(T_t(\alpha_t)|Y_t), \\
    P_{t+1} &= \text{Var}(\alpha_{t+1}|Y_t) \\
    &= \text{Var}(T_t(\alpha_t)|Y_t) + H_t.
\end{align*}

(2.36)

At this point, the chief obstacle is the evaluation of the expectations of non-linear functions of random variables. If we have a method to calculate \( \mathbb{E}(f(\alpha_t)|Y_{t-1}) \) for some generic transformation \( f(\cdot) \) where \( \mathbb{E}(\alpha_t|Y_{t-1}) = a_t \) is known, we can evaluate (2.35) and (2.34). (Note that the variances and cross covariances in (2.35) are also expectations of non-linear transformations of \( \alpha_t|Y_{t-1} \)). Subsequently if we can calculate \( \mathbb{E}(g(\alpha_t)|Y_t) \) for some \( g(\cdot) \) when \( \mathbb{E}(\alpha_t|Y_t) = a_{t|t} \) is known, we can evaluate (2.36), and thus complete the recursion. One suitable method is the unscented transformation by Julier et al. (1995).

2.4.1 Unscented transformation

Suppose we have a non-linear function \( y = f(x) \) of a random variable \( x \) with known mean \( \bar{x} \) and variance \( \sigma^2 \) and we wish to find the mean and variance of the transformed variable \( y \), assuming that they exist. For complicated functions, the exact solutions are often intractable, and approximations are often sought in grid or simulation based numerical integration methods. In the most basic simulation procedure, a large number of samples are drawn from the distribution of \( x \), and \( \mathbb{E}(y) \) is approximated by the sample average of of the transformed sample points. In the unscented transformation, the expectation is approximated by the weighted average of the transformation of a small set of deterministically chosen sigma points \( \{x_i\} \) with associated weights \( \{w_i\} \), that is

\[ \mathbb{E}(y) \approx \sum_i w_i f(x_i) = \bar{y}. \]  

(2.37)
Other statistics are estimated in a similar fashion, e.g.,

\[ \text{Var}(y) = \mathbb{E}(y - \mathbb{E}(y))^2 \approx \sum_i w_i (f(x_i) - \bar{y})^2. \] (2.38)

In order to choose suitable sigma points \( x_i \) and associated weights \( w_i \) we considering the Taylor representations of \( f(x_i) \) and \( f(x) \). Expanding \( f(x_i) \) around \( \bar{x} \) and substituting in (2.37) yields

\[
\bar{y} = \sum_{j=0}^{\infty} \sum_i w_i \frac{(x_i - \bar{x})^j}{j!} \left. \frac{d^j f(z)}{dz^j} \right|_{z=\bar{x}}
\] (2.39)

while the expectation of the Taylor expansion of \( f(x) \) is given by

\[
\mathbb{E}[f(x)] = \sum_{j=0}^{\infty} \mathbb{E}(x - \bar{x})^j \left. \frac{d^j f(z)}{dz^j} \right|_{z=\bar{x}}
\] (2.40)

This expectation can be approximated by (2.39) by equalising the Taylor series term by term, which yields the series of conditions on the sigma points and weights

\[
\sum_i w_i (x_i - \bar{x})^j = \mathbb{E}(x - \bar{x})^j, \quad j = 0, 1, 2, \ldots
\] (2.41)

The first three of these constraints are given by

\[
\begin{align*}
\sum_i w_i &= 1, \quad (2.42a) \\
\sum_i w_i x_i &= \bar{x}, \quad (2.42b) \\
\sum_i w_i (x_i - \bar{x})^2 &= \sigma^2. \quad (2.42c)
\end{align*}
\]

These restrictions do not uniquely determine the sigma points and weights, but information of higher moments of \( x \) can be used to impose more constraints. For instance, when Gaussianity is assumed for \( x \), the addition of the condition imposed by the fourth central moment and symmetry around the mean fixes the points and weights. The result is a method in which, with a deterministic “sample” of merely three points from \( x \), the true expectation of \( y = f(x) \) can be calculated exactly for arbitrary differentiable transformations \( f(\cdot) \), without requiring the calculation of any derivatives for the first four orders of the Taylor expansion (and the first two orders of the variance estimate). Although higher order terms are not matched, their influence are expected to decline rapidly due to the inverse factorial weighting. The sigma points and weights are chosen based on the mean and variance of \( x \) and do not depend on \( f(\cdot) \).

For simplicity of illustration, the development above pertains scalar random variables. The same principle applies for vector valued random variables, in which case
multi-variate Taylor approximations are used in (2.39) and (2.40). The first two conditions in (2.42) remain the same, while the matching of the second moment becomes

$$\sum_i w_i (x_i - \bar{x})(x_i - \bar{x})' = P_{xx},$$

(2.43)

where $P_{xx}$ is the variance matrix of $x$.

Although the procedure bears some resemblance to Monte-Carlo estimation, the number of points used in the unscented transformation is pre-determined and typically very small. To some degree it can be helpful to interpret the weights as the probability density of a random variable with the sigma points as the outcome space, and which mimics the reference variable. A limit to this interpretation is that some of the weights $w_j$ can in general be negative.

### 2.4.2 Sigma point selection

There are many ways to choose the sigma points and weights subject to the constraints (2.42). A well chosen set can match higher order terms of the Taylor approximations by incorporating knowledge of the distribution of $x$. For $n$-dimensional random vector $x$ with mean $\bar{x}$ and variance $P_{xx}$, Julier and Uhlmann (1997) suggest choosing a set of $2n + 1$ sigma points of the form

$$x_0 = \bar{x}, \quad x_i = \bar{x} + \lambda \sqrt{P_{xx,i}}, \quad x_{i+n} = \bar{x} - \lambda \sqrt{P_{xx,i}},$$

(2.44)

for $i = 1, \ldots, n$, where the scalar $\lambda$ is a free parameter and $\sqrt{P_{xx,i}}$ is the $i$th column of a matrix square root of $P_{xx}$, which can be obtained by for example a Cholesky decomposition of the matrix. The constant $\lambda$ is determined by the constraints. It is simple to verify that the conditions of (2.42),(2.43) are satisfied for $w_i = w_{i+n} = (1 - w_0)/2n$ and $\lambda^2 = n/(1 - w_0)$. If $x$ is a Gaussian vector with a diagonal covariance matrix, we can use the knowledge of the kurtosis to impose an additional restriction, with which we can solve $w_0 = 1 - n/3$. Thus, the sigma points and weights for Gaussian vectors with diagonal covariance matrix are given by

$$x_0 = \bar{x}, \quad w_0 = 1 - n/3$$

$$x_i = \bar{x} + \sqrt{3} \sqrt{P_{xx,i}}, \quad w_i = 1/6$$

$$x_i = \bar{x} - \sqrt{3} \sqrt{P_{xx,i}}, \quad w_i = 1/6.$$ 

(2.45)

The unscented transformation based on these points and weights gives an approximation of the mean which is exact up to the second order, and up to the fourth order in its entries on the diagonal. The covariance estimate is accurate up to the second order.

In the engineering literature some effort have been undertaken to devise sigma-point selection schemes using an economic amount of points. Tenne and Singh (2003) showed
2.4. THE UNSCENTED KALMAN FILTER

how accuracy of the approximation can be increased by solving increasingly larger sets of moment matching equations. Their examples demonstrate solutions for univariate Gaussian input variables with five and seven sigma points, providing accuracy up to the eight and twelfth order moments respectively. A drawback of the method is that finding solutions for higher order approximations becomes increasingly complicated.

2.4.3 Unscented filtering summary, smoothing and parameter estimation

Combining the results of the previous subsections, the unscented filter can be implemented as follows.

1. Calculate sigma-points and weights $x_i, w_i$ based on $a_t, P_t$.

2. Use the unscented transformation to estimate $\tilde{y}_t = \sum_i w_i z_t(x_i)$.

3. Calculate the gain matrix $K_t = P_{ay,t} P_{yy,t}^{-1}$ using the unscented transformation for the estimates $P_{ay,t} = \sum_i w_i (x_i - a_t)(z_t(x_i) - \tilde{y}_t)'$ and $P_{yy,t} = \sum_i w_i (z_t(x_i) - \tilde{y}_t)(z_t(x_i) - \tilde{y}_t)' + G_t$.

4. Estimate $a_{t|t} = a_t + K_t(y_t - \tilde{y}_t)$ and $P_{t|t} = P_t - K_t P_{yy,t} K_t'$.

5. Calculate new sigma-points and weights $\tilde{x}_i, \tilde{w}_i$ based on $a_{t|t}, P_{t|t}$.

6. Estimate using unscented transformations $a_{t+1} = \sum_i \tilde{w}_i T_t(\tilde{x}_i)$ and $P_{t+1} = \sum_i \tilde{w}_i (T_t(\tilde{x}_i) - a_{t+1})(T_t(\tilde{x}_i) - a_{t+1})' + H_t$.

A non-linear smoother using the unscented transformation can be derived from the linear Gaussian smoother in the same manner as the unscented filter. Using the notation as before, with

$$E(\alpha_t|Y_t) = a_{t|t}, \quad E(\alpha_{t+1}|Y_t) = a_{t+1}, \quad Var(\alpha_t|Y_t) = P_{t|t}, \quad Var(\alpha_{t+1}|Y_t) = P_{t+1}, \quad Cov(\alpha_t, \alpha_{t+1}|Y_t) = M_t,$$

we notice that the derivation of the linear smoother (2.13) does not depend on the linearity of the transition an the measurement equations. Hence, if we assume or approximate the relevant conditional distributions by Gaussians in each iteration step, we can use the formula

$$\hat{\alpha}_t = a_{t|t} + M_t P_{t+1}^{-1}(\hat{\alpha}_{t+1} - a_{t+1}),$$

(2.49)
where $a_{t+1}, a_{t|t}, P_{t|t}$ and $v_{t+1}$ are calculated using the unscented filter, and the covariance is approximated using an unscented transformation according to

$$M_t = \sum_{i=0}^{2m} w_i (\tilde{x}_{t,i} - a_{t|t})(T_t(\tilde{x}_{t,i}) - a_{t+1})'. \tag{2.50}$$

Parameter estimation using the unscented filter can be performed in a similar manner as in the standard and extended filter, using the prediction error decomposition (2.18) with the approximate prediction errors $v_t = y_t - \bar{y}_t$ and variances $F_t = P_{yy,t}$ obtained as described earlier in this subsection. As the model approximation from the unscented filter is generally superior to that of the extended filter, it can be expected to be more accurate than the similar procedure performed with the prediction errors from the extended filter.

### 2.5 State estimation error comparisons

In order to provide some guidance to the comparative performance of between the extended and the unscented filters and smoothers, we provide simulation results from a simple non-linear model. Observations are generated as non-linear transformations of a state which follows a basic stationary AR(1) process

$$\begin{align*}
\alpha_{t+1} &= 0.95 \alpha_t + \eta_t, \\
y_t &= Z(\alpha_t) + \epsilon_t,
\end{align*} \quad \eta_t \sim \text{NID}(0, 0.1), \quad \epsilon_t \sim \text{NID}(0, 0.1), \quad t = 1, \ldots, 100. \tag{2.51}
$$

Table 2.1 shows the average over 10000 simulation runs of the mean squared state estimation error (MSE) and the mean absolute state estimation error (MAD). The state estimates are calculated by the extended filter (EKF), extended smoother (EKS), unscented filter (UKF) and unscented smoother (UKS). The table shows that the unscented filters and smoother generally yield more accurate estimates than their extended counterparts. For the exponential and logarithmic transformation, the differences are small to negligible. However, the improvement for the sine and polynomials are considerable. Generally the smoother gains more from the unscented approximation than the filter.

### 2.6 Alternative methods

Many other non-simulation based filtering algorithms have appeared in especially the engineering literature. We mention a few here.
### Table 2.1: Mean squared and absolute state estimation errors of extended and unscented Kalman filters and smoothers.

<table>
<thead>
<tr>
<th>Function</th>
<th>MSE</th>
<th></th>
<th>MSE</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EKF</td>
<td>EKS</td>
<td>UKF</td>
<td>UKS</td>
</tr>
<tr>
<td>(Z(\alpha_t))</td>
<td>0.1001</td>
<td>0.0758</td>
<td>0.0937</td>
<td>0.0667</td>
</tr>
<tr>
<td>(\exp(\alpha_t))</td>
<td>0.4204</td>
<td>0.2990</td>
<td>0.4198</td>
<td>0.2951</td>
</tr>
<tr>
<td>(\log(\alpha_t + 6))</td>
<td>0.2896</td>
<td>0.2687</td>
<td>0.2261</td>
<td>0.2014</td>
</tr>
<tr>
<td>(\sin(\alpha_t))</td>
<td>0.9187</td>
<td>0.9186</td>
<td>0.4008</td>
<td>0.2968</td>
</tr>
<tr>
<td>(\alpha_t^3/10)</td>
<td>0.9229</td>
<td>0.9229</td>
<td>0.3142</td>
<td>0.2400</td>
</tr>
<tr>
<td>(\alpha_t^5/10)</td>
<td>0.9384</td>
<td>0.9384</td>
<td>0.4117</td>
<td>0.4072</td>
</tr>
</tbody>
</table>

As the extended filter relies on a first order approximation of the non-linearities of the model, an obvious extension is to expand the approximation order. Classic references on higher order extended filters are Jazwinski (1970) and Maybeck (1982). In practice, higher order filters are quite cumbersome to implement, requiring the calculation of matrices and tensors of derivatives. As such, they have generally found little uptake outside some specialised applications.

Another method of improving the extended filter is to seek improvements in the linearisation points. Running the standard single pass extended filter, the Taylor approximations are expanded around the most recently available state estimates (\(a_{it}\) and \(a_{it+1}\) for the transition and observation equations respectively). It can be reasonably surmised that expansions around a more accurate state estimate can yield an improvement. This thought underlies the iterated extended Kalman filter. In this algorithm, after the estimating the state from the extended filter and smoother, the filter and smoother are re-run with linearisations around the smoothed state estimates. This procedure yields a new set of smoothed estimates, and is then iterated until convergence. Like the higher order extended filters, this may provide a worthwhile improvement for specific problems, although few theoretical guidelines are available. Again, Jazwinski provides a extensive descriptions and examples.

A separate and more sophisticated approach is the Gaussian sum filter (Sorensen and Alspach, 1971), (Alspach and Sorenson, 1972). Underlying this method is the approximation of the filtering density by a weighted sum of Gaussian densities \(\varphi(\cdot|\mu, \sigma)\)

\[
p(\alpha_t|Y_t) \approx \sum_{i=1}^{m} w_{it} \varphi(\alpha_{it}|\mu_{it}, \sigma_{it}).
\]

In its implementation, the parameters of the components in the sum are calculated.
by running a bank of $m$ extended Kalman filters in parallel. For details, we refer to Anderson and Moore (1979, Chap. 8). Recently, this approach has been improved by using unscented filters instead of extended filters in the summation terms, see Simandl and Dunik (2006).

Finally, traditional grid based numerical integration methods such as Legendre or Hermite quadratures (Stoer and Bulirsch, 1980) have been applied to non-linear filtering problems, see Kitagawa and Gersch (1996). The classical problem of numerical integrations is the explosive computational burden in high-dimensional problems. Although problem specific simplifications can sometimes be devised for certain state space models which make numerical integration feasible, simulation or Monte Carlo based methods have seen far wider application in the past few decades.
Chapter 3

Monte Carlo Filtering Algorithms

3.1 Introduction

The full Bayesian estimation problem in the general state space model is essentially a problem of solving high dimensional integrals. Explicit closed form solutions are only available under some quite restrictive assumptions, while classical grid based numerical integration methods are problematic as the number of calculations tends to rise exponentially with the dimension of the state. In the previous chapter, the extended and unscented filter were discussed as approximate solutions, which rely on replacing some aspects of the non-linear model with simplified structures. By contrast in the simulation or Monte Carlo approach, approximations are performed on the estimators, rather than the model. An important advantage of simulation methods is that increasing the accuracy of the approximation is in principle as straightforward as letting computer programs run longer.

Monte Carlo integration has been studied well before the development of the Kalman filter. Many of the key mathematical results were developed by physicists and mathematicians at the Los Almos National Laboratory in the 1940s and 1950s. However, only since the 1980s, with the wide availability of cheap computing power, did applications of the Monte Carlo method to realistic state space models become practical. This initiated intense research activity, leading to a large number of algorithms and publications. While we restrict the detailed descriptions in this chapter to importance sampling based methods, many other simulation techniques for estimating state space models have been devised. Especially noteworthy are Markov chain Monte Carlo (MCMC) methods which have greatly influenced research in Bayesian analysis in general, see Gamerman (1998), Andrieu et al. (1998) and Gilks et al. (1980). Another method based on classical Monte Carlo techniques is rejection sampling, which is discussed in the state space context by Tanizaki and Mariano (1998).
3.2 Importance sampling

Importance sampling is a general technique for estimating probability distributions when only samples from some other related distribution can be generated. It was introduced by Marshall (1956) and appeared first in the econometrics literature in Kloek and van Dijk (1978). Specializations to non-linear state space models as described in this chapter were developed by Shephard and Pitt (1997) and Durbin and Koopman (1997). A general statistical description can be found in Ripley (1987).

The basis of Monte Carlo integration methods is that integrals of the form

\[ I = \int x(\alpha) p(\alpha) \, d\alpha \]  

(3.1)

where \( p(\cdot) \geq 0 \) and \( \int p(\alpha) \, d\alpha = 1 \) can be interpreted as the expectation of a function \( x(\cdot) \) of a random variable which has probability density \( p(\cdot) \). The law of large numbers asserts the convergence of (functions of) the sample average of observations drawn from the density to (functions of) its expectation. This justifies approximating the integral \( I \) by drawing \( N \) random samples \( \alpha^{(i)} \) from \( p(\cdot) \) and calculating the average

\[ \frac{1}{N} \sum_{i=1}^{N} x(\alpha^{(i)}). \]  

(3.2)

Often there is no known method to draw samples from \( p(\cdot) \). However, if an importance density \( g(\cdot) \) is available from which samples can be generated, and \( g(\cdot) \) is a reasonable approximation of \( p(\cdot) \), the integral can be calculated as

\[ I = \int x(\alpha) \frac{p(\alpha)}{g(\alpha)} g(\alpha) \, d\alpha. \]  

(3.3)

The approximation of \( x(\alpha) \) is then performed by drawing samples \( \alpha^{(i)} \) from the distribution associated with \( g(\cdot) \) and calculating the average

\[ \frac{1}{N} \sum_{i=1}^{N} x(\alpha^{(i)}) w(\alpha^{(i)}), \]  

(3.4)

where

\[ w(\alpha^{(i)}) = \frac{p(\alpha^{(i)})}{g(\alpha^{(i)})} \]  

(3.5)

are known as the importance weights.

In state space models, we seek to estimate the expectation and other statistics of the state \( \alpha \) conditional on observations \( y = (y_1', \ldots, y_n') \). In non-linear models, it is generally unknown how to sample directly from \( p(\alpha|y) \). The main problem when we wish to use importance sampling is to find an approximating density \( g(\alpha|y) \), which resembles \( p(\alpha|y) \) and from which we can draw samples. The estimates can then be calculated as outlined above.
3.2 IMPORTANCE SAMPLING

3.2.1 Importance densities from linear Gaussian state space approximation

The non-linear model (1.1) can be approximated using a first order Taylor expansion

\[
\begin{align*}
\alpha_{t+1} &\approx T_t(\hat{\alpha}_t) + \dot{T}_t \cdot (\alpha_t - \hat{\alpha}_t) + \eta_t, \\
y_t &\approx Z_t(\hat{\alpha}_t) + \dot{Z}_t \cdot (\alpha_t - \hat{\alpha}_t) + \epsilon_t, \\
\dot{T}_t &= \frac{\partial T_t(x)}{\partial x} \bigg|_{x=\hat{\alpha}_t}, \\
\dot{Z}_t &= \frac{\partial Z_t(x)}{\partial x} \bigg|_{x=\hat{\alpha}_t}.
\end{align*}
\] (3.6a, 3.6b, 3.6c)

This is a modified linear state space model as in (2.25) with

\[
\begin{align*}
d_t &= T_t(\hat{\alpha}_t) - \dot{T}_t \hat{\alpha}_t, \\
c_t &= Z_t(\hat{\alpha}_t) - \dot{Z}_t \hat{\alpha}_t
\end{align*}
\] (3.7, 3.8)

Starting from some initial \(\hat{\alpha}_t\) (which may for instance be obtained from an extended Kalman smoother), the state of the linearised model can be estimated using a linear Kalman smoother and iterated until the state convergences to \(\hat{\alpha}_t = \hat{\alpha}_t\). Substituting \(\hat{\alpha}_t\) in (3.6) yields a linear Gaussian approximating model, which serves as the basis for the importance density.

To estimate the conditional mean, or some other function \(x(\alpha|y)\) of the conditional state, note that

\[
E(x(\alpha|y)) = \int x(\alpha) \frac{p(\alpha|y)}{g(\alpha|y)} g(\alpha|y) d\alpha
\]

\[
= E_g \left[ x(\alpha) \frac{p(\alpha|y)}{g(\alpha|y)} \right],
\] (3.9)

\[
= \frac{g(y)}{p(y)} E_g \left[ x(\alpha) \frac{p(\alpha, y)}{g(\alpha, y)} \right],
\]

where \(E_g[\cdot]\) is the conditional expectation under density \(g(\cdot)\). Noting that for the function \(x(\alpha) = 1\) we have

\[
1 = \frac{g(y)}{p(y)} E_g \left[ \frac{p(\alpha, y)}{g(\alpha, y)} \right],
\] (3.10)

we can divide (3.9) by (3.10) to obtain

\[
E(x(\alpha|y)) = \frac{E_g[x(\alpha)w(\alpha)]}{E_g[w(\alpha)]}, \quad w(\alpha) = \frac{p(\alpha, y)}{g(\alpha, y)}.
\] (3.11)

Hence, we can estimate \(E(x(\alpha|y))\) by drawing \(N\) samples \(\alpha^{(i)}\) from the approximating model density \(g(\alpha|y)\) and calculating

\[
\sum_i \frac{x(\alpha^{(i)}) \cdot w(\alpha^{(i)})}{\sum_i w(\alpha^{(i)})}.
\] (3.12)
The ratios of the true model density and the approximating density

\[
w_i = \frac{p_{\theta}(\alpha^{(i)}, y)}{g_{\theta}(\alpha^{(i)}, y)}
\]

are known as importance weights. The joint densities \( p(\alpha, y) \) and \( g(\alpha, y) \) required to calculate the weights follow directly from the state space model definitions:

\[
p(\alpha, y) = p(\alpha_1) \prod_{t=1}^{n} p(y_t|\alpha_t) p(\alpha_{t+1}|\alpha_t)
\]

\[
= p(\alpha_1) \prod_{t=1}^{n} p_{\epsilon_t}(y_t - Z_t(\alpha_t)) p_{\eta}(\alpha_{t+1} - T_t(\alpha_t))
\]

and

\[
g(\alpha, y) = g(\alpha_1) \prod_{t=1}^{n} g_{\epsilon_t}(y_t - \hat{Z}_t(\alpha_t) - d_t) g_{\eta}(\alpha_{t+1} - \hat{T}_t(\alpha_t) - c_t)
\]

Finally, we require an algorithm to draw samples from the conditional state \( \alpha|y \) for linear Gaussian models in order to produce samples from \( g(\alpha|y) \). Such algorithms are termed simulation smoothers, and have been developed by Frühwirth-Schnatter (1994), Carter and Kohn (1994), de Jong and Shephard (1995) and Durbin and Koopman (2002). The procedure of Durbin and Koopman appears to be the most straightforward to implement, and can be summarised as follows. Given a linear Gaussian state space model (2.25) and observations \( y \)

1. Generate unconditional draws \( \alpha^{+(i)} \) and \( y^{+(i)} \) from the state equation (2.25a) and observation equation (2.25b).

2. Apply the standard Kalman smoother to both the true observations \( y \) and the simulated observations \( y^{+(i)} \), yielding \( \hat{\alpha} \) and \( \hat{\alpha}^{+(i)} \) respectively.

3. Calculate \( \alpha^{(i)} = \hat{\alpha} + \hat{\alpha}^{+(i)} - \alpha^{+(i)} \).

The resulting samples \( \alpha^{(i)} \) are distributed according to \( \alpha|y \).

### 3.2.2 Parameter estimation using importance sampling

The importance sampling method is very useful to produce estimates of the likelihood in non-linear models. In state space models with parameter vector \( \theta \), the likelihood function is given by

\[
L(\theta) = p_{\theta}(y) = \int p_{\theta}(\alpha, y) d\alpha,
\]

(3.16)
3.2. IMPORTANCE SAMPLING

where $p_\theta(\cdot)$ the joint density of the data and the state. (Note that in this section, we subscript probability densities by parameters, rather than the random variable which they represent.) In non-linear state space models the likelihood or marginal density of the observations $p_\theta(y)$ is usually unknown. The joint density $p_\theta(\alpha, y)$ can be obtained from the state space model definition by factoring it as

$$
p_\theta(\alpha, y) = \prod_{t=1}^{n} p_\theta(y_t|\alpha_t)p_\theta(\alpha_t|\alpha_{t-1}),$$  

(3.17)

with $p_\theta(\alpha_1|\alpha_0) = p_\theta(\alpha_1)$. The individual factors $p_\theta(y_t|\alpha_t)$ and $p_\theta(\alpha_t|\alpha_{t-1})$ are defined by the state space observation and transition models. Although the likelihood can in principle be calculated by integrating the joint density over the state numerically, this is normally not feasible due to the high dimensionality of $\alpha$ and the fact that the most important application (numerical likelihood maximisation) usually requires the integration to be performed many times. Using Monte Carlo integration principles $L(\theta) = \int p_\theta(y|\alpha)p_\theta(\alpha) \, d\alpha$ can be estimated by drawing $\alpha^{(i)}$ according to the distribution $p_\theta(\alpha)$ and calculating the mean of $p_\theta(y|\alpha^{(i)})$. However, this naive Monte Carlo likelihood is extremely inefficient. The majority of the generated $\alpha^{(i)}$s will diverge very much far from the true $\alpha$ and therefore contribute a negligible amount to the likelihood. In practice this means that a prohibitive number of draws of $\alpha^{(i)}$ are required in order to obtain an accurate estimate. Ideally, $\alpha^{(i)}$ should be simulated from the true density conditional on the observations, i.e., from $p_\theta(\alpha|y)$, but for non-linear models this is usually not possible. Using importance sampling with an approximating linear Gaussian model as in the previous section, a feasible Monte Carlo likelihood estimate can be produced.

First, rewrite the likelihood as

$$L(\theta) = \int p_\theta(\alpha, y) \, d\alpha = \int \frac{p_\theta(\alpha, y)}{g_\theta(\alpha|y)} g_\theta(\alpha|y) \, d\alpha = g_\theta(y) \int \frac{p_\theta(\alpha, y)}{g_\theta(\alpha, y)} g_\theta(\alpha|y) \, d\alpha.$$  

(3.18)

The first factor $g_\theta(y)$ is the likelihood of the approximating model; the integral is the expectation of $p_\theta(\alpha, y)/g_\theta(\alpha, y)$ under the distribution of $g_\theta(\alpha|y)$. Hence, the log-likelihood is estimated by

$$\log \hat{L}(\theta) = \log L_g(\theta) + \log \bar{w},$$  

(3.19)

where

$$\bar{w} = \frac{1}{N} \sum_{i=1}^{N} \frac{p_\theta(\alpha^{(i)}, y)}{g_\theta(\alpha^{(i)}, y)} = \frac{1}{N} \sum_{i=1}^{N} w(\alpha^{(i)}),$$  

(3.20)
where \( L_\theta(y) = g_\theta(y) \) is the likelihood from the approximating model and \( \alpha^{(i)} \) are drawn from \( g_\theta(\alpha|y) \) using the approximation method and simulation smoothing algorithm as described in the previous subsection.

### 3.2.3 Diagnostic testing for importance sampling

In several texts on importance sampling, it is noted that for numerically stable estimates, the variation in the importance weights \( w(\alpha^{(i)}) \) should be bounded (Geweke, 1989; Monfort and van Dijk, 1995). Simple diagnostics include investigating the size of the largest drawn weights, and plotting the empirical distribution of the weights.

The importance sampling techniques employed for the estimation of the likelihood function can be justified on the basis of the central limit theorem (CLT). Although the consistency property \( \hat{L}(\psi) \rightarrow L(\psi) \) as \( N \rightarrow \infty \) always holds due to the law of large numbers, the CLT is generally not valid when the second moment of the importance ratio \( p_\theta(\alpha, y)/g_\theta(\alpha, y) \) does not exists.

Koopman et al. (2009) suggest to carry out testing procedures based on the existence of the variance. These test statistics arise from extreme value theory. For example, importance weights that are larger than a certain threshold value are assumed to come from a generalised Pareto distribution. After calculating a large number of weights, the existence of the second moment can then be formally tested using tests on the parameters of the Pareto distribution.

To carry out the test, first we take all weights larger than a fixed threshold value \( u \). For increasing \( u \), the limiting distribution of the weights minus \( u \) is generalised Pareto, with has the density function given by

\[
f(w) = \frac{1}{\beta}\left(1 + \frac{w}{\beta}\right)^{-\frac{1}{\xi} - 1}.
\] (3.21)

The scale and shape parameters \( \beta \) and \( \xi \) can be estimated from generated weights by maximum likelihood. Since the Pareto distribution has \( 1/\xi \) finite moments, testing the existence of the variance is performed by testing \( H_0 : \xi = 1/2 \) against \( H_1 : \xi > 1/2 \), using classical likelihood based testing procedures.

Under the null hypothesis of a finite second moment, the Wald and Lagrange Multiplier (LM) tests have a standard Gaussian distribution, while the likelihood ratio (LR) test is distributed as \( 0.5(\chi^2_0 + \chi^2_1) \). An alternative test based on the largest order statistics of the weights was developed by Monahan (1993). This statistic has a negative value under the null.
3.3 Particle filtering

Particle filtering is a simulation technique which has quickly gained in popularity in the past 15 years, especially the engineering disciplines. It is amongst the most complete solutions to filtering problems in the general non-linear model currently available. The particle filter attempts to track the evolution of the entire filtering and prediction density of the Bayesian solutions (1.10–1.9) by representing them by simulated samples drawn from the distributions. This contrasts to the methods of the previous chapter, which only attempts to keep track of the means and variance of the state, imposing simplifying assumptions on the model in each step. In physical terms, methods such as the extended and unscented filter emphasise the estimation of the position of the state, while particle filtering concerns estimating the entire probability distribution.

3.3.1 Generic particle filter

A generic particle filter is essentially a sequential Monte Carlo method using importance sampling. The approximating importance density is constructed at each time step, in contrast to the the global importance sampling techniques described before, in which the entire approximating model is constructed before beginning the sampling stage.

A general description is as follows. In step \( t \) we start with a set of \( N \) particles \( \tilde{\alpha}_t^{(i)} \) and weights \( \tilde{w}_t^{(i)} \), which are samples representing the distribution \( p(\alpha_t|Y_{t-1}) \). Initially the particles receive uniform weights \( \tilde{w}_t^{(i)} = N^{-1} \) for all \( i \), so that the set of particles and weights are \( \{\tilde{\alpha}_t^{(i)}, N^{-1}\} \). For each particle, new importance weights are computed based on the importance density \( q(\cdot) \) and observation \( y_t \). This results in a weighted \( \{\tilde{\alpha}_t^{(i)}, \tilde{w}_t^{(i)}\} \) which approximates the filter density \( p(\alpha_t|Y_t) \). The weights need to be renormalised before proceeding to the next iteration step. However, it was shown in Kong et al. (1994) that the variance of the importance weights increase over time, which leads to the a degeneracy problem: after a few iterations, most of the weights are close to zero. This implies that many particles will barely contribute to the distribution estimate after a few steps. In particle filtering, the key to solving this problem is a resampling step, which selects the particles “fit for survival” according to some criterion, removing the particles with low importance weights, and redraw new particles to proceed. These are transformed in the prediction step to form a new set of unweighted particles \( \{\tilde{\alpha}_{t+1}^{(i)}, N^{-1}\} \) for the next iteration. The estimate of the filtered distribution in step \( t \) is the empirical distribution given by

\[
\hat{p}(\alpha_t|y_{1:t}) = \sum_{i=1}^{N} \tilde{w}_t^{(i)} \delta(\alpha_t - \tilde{\alpha}_t^{(i)}),
\]

where \( \delta(\cdot) \) is the Dirac delta function, and expectations and other statistics \( x(\alpha_t|y_{1:t}) \)
are approximated by

\[ \hat{x} = \sum_{i=1}^{N} x(\alpha_t^{(i)}) \hat{w}_t^{(i)}. \] (3.23)

The main practical issues in particle filtering are the selection of the proposal density, and the resampling procedure. The proposal densities we consider have the natural factorised form

\[ q(\alpha_{1:t}|y_{1:t}) = q(\alpha_1) \prod_{t=1}^{n} q(\alpha_t|\alpha_{1:t-1}, y_{1:t}), \] (3.24)

although densities which take future observations into account have also been proposed, e.g., Pitt and Shephard (1999). In order to mitigate the degeneracy problem, a natural choice is a proposal density which minimises the variance of the importance weights. As shown by Doucet et al. (1999), the theoretically optimal proposal using this criterion is the density

\[ q(\alpha_t|\alpha_{1:t-1}, y_{1:t}) = p(\alpha_t|\alpha_{1:t-1}, y_{1:t}), \] (3.25)

defined by the true non-linear model. However, this distribution is generally unavailable for evaluation and sampling (since this is the chief motivation to use simulation methods).

A rather simplistic proposal choice follows immediately from the model definition is

\[ q(\alpha_1|\alpha_{1:t-1}, y_{1:t}) = p(\alpha_t|x_{t-1}). \] (3.26)

Although not particularly sophisticated, this choice has been very popular in the literature (Kong et al., 1994; Liu and Chen, 1995; Gordon et al., 1993; Kitagawa and Gersch, 1996).

### 3.3.2 Extended and unscented particle filters

Heuristically, the particle degeneracy problem is most effectively dealt with by moving the particles to the region of high likelihood of the most current observations. One method to accomplish this is by means of the extended Kalman filter. For each particle, an extended filter is run, producing a particle specific state and variance prediction. The particle specific proposal density is given by the Gaussian density of the linearised model, and a new particle draw can be generated using the standard linear sampling tools. This has been shown to improve upon the simple proposal (3.26) by Doucet (1998).

An obvious further extension is to replace the extended filter by the unscented filter. The resulting algorithm has been called the unscented particle filter. It has been developed and studied in Merwe et al. (2000), Haykin (2001) where it demonstrated superior
performance compared to the extended filter based proposal for a diverse selection of examples.
CHAPTER 3. MONTE CARLO FILTERING ALGORITHMS
Chapter 4

Unobserved Components Cycles with Smooth-Transition Parameters

Abstract

To gain insights in the current status of the economy, macro economic time series are often decomposed into trend, cycle and irregular components. The dynamic properties of trend and cycle variations in economic time series are not necessarily constant over time in a long sample period. In the context of model-based trend-cycle decompositions, more flexible model specifications may need to be considered to account for the time variation of parameters that determine the dynamic properties of a time series. In this chapter we empirically investigate the effectiveness of smooth transition mechanisms for parameters that drive the cycle components. Specifically, we investigate whether the business cycle volatility and period remain constant over time and give likelihood based information criteria for this purpose. These developments are illustrated by analysing four different time series from the U.S. economy (GDP, investment, unemployment and industrial production).

4.1 Introduction

Macroeconomic time series are usually subject to cyclical dynamics that can be relevant for economic policy analysis. For example, the cyclical deviations in gross domestic product time series from its long term trend are usually associated with the business cycle. Since the business cycle is unobserved, we require methods to extract business cycle indicators from macroeconomic time series. Various approaches exist for the signal extraction of a cycle from a univariate time series. It is common practice to apply non-parametric filters such as the ones of Hodrick and Prescott (1997) and Baxter and
King (1999). The filter is formulated in terms of a symmetric weighting kernel and provides consistent estimates of trends and cycles in the middle of the time series while adjustments are required at the beginning and end of the time series. More recently the importance of frequency domain properties has been emphasised. For example, the Baxter-King filter is referred to as a bandpass filter since it aims to extract the cyclical dynamics that are associated with so-called business cycle frequencies $2\pi/\omega$, where the period $\omega$ ranges from 1.5 to 8 years. The effectiveness of such a filter is considered in the frequency domain where the gain function should only affect the business cycle frequencies. It is often observed that many business cycle filters still affect higher frequencies resulting in noisy cycle indicators. Christiano and Fitzgerald (2003) and Azevedo (2003) have considered “near optimal” bandpass filters that retain the pre-specified business cycle frequencies almost perfectly. Further, they deal with the endpoint problems of non-parametric filters.

Non-parametric filters work on the basis of a clearly defined aim of extracting the sum of frequency components of a time series that are related to the business cycle. This clarity can also be viewed as a problem with non-parametric filtering, since it does not take into account the dynamic properties of a time series. In the extreme case, when the time series is generated by a white noise process, the non-parametric filter still extracts a cyclical component from the time series. Since the statistical properties of a given time series are not known, this calls for the consideration of a business cycle indicator that is the result of an analysis based on a statistical model. Three of the advantages of the model-based approach are that it is able to predict future values, to construct confidence intervals and to test parameters for significance. Examples of model-based approaches to business cycles are given by Beveridge and Nelson (1981), Clark (1987) and Harvey and Jaeger (1993).

The so-called Beveridge-Nelson decomposition is based on the modelling of a univariate time series by an autoregressive integrated moving average (ARIMA) model. The lag polynomials for the autoregressive part and the moving average part with estimated coefficients can be decomposed into a sum of two ARIMA components; one is based on a unit root autoregressive polynomial (the trend) and the other is a stationary component (the cycle). The former polynomial enables the extraction of the trend component while the cycle component is taken as the deviations of the observations from the estimated trend. This approach is known as the canonical decomposition and is also used in the context of seasonal adjustment; see Burman (1980) and Maravall (1985).

The model-based decompositions established by Clark (1987) and Harvey and Jaeger (1993) are based on unobserved components time series models with trend and cycle factors. The parameters of the model are estimated by the method of maximum likelihood.
for which state space methods can be adopted to evaluate the log-likelihood function. Kalman filtering and smoothing methods can also be used for the signal extraction of the cycle component. The estimated cycle factor depends on the formulation of the unobserved components present in the model. For example, Clark’s model specifies the cycle component as an autoregressive process while the Harvey-Jaeger model has the cycle as a stationary and stochastic cyclical process based on a cosine wave.

In this chapter we adopt the latter modelling strategy for measuring the business cycle from U.S. macroeconomic time series. In particular we focus on the stability of parameters that determine the cyclical component. These parameters may change over time or may vary depending on periods of expansion and contraction. An interesting parameter of the cycle component in this respect is the variance of the stochastic shocks that enter the model for the cycle component. This parameter determines the variability or volatility of the business cycle. For example, Stock and Watson (2003) argue that the volatility of economic activity in most G7 economies has moderated since the 1960s. They adopt structural vector autoregressive models in their analyses. Volatility breaks in G7 macroeconomic series are also overwhelmingly found in the study of van Dijk et al. (2002). However, such findings have not yet been established within the trend-cycle decomposition modelling framework.

A treatment of cycle parameter variation in unobserved component models has not been considered before with the exception of Koopman et al. (2005), where different variances are estimated for three different time intervals. In this chapter we extend the flexibility of the parametrisation by considering variations in the persistence and the period of the cycle and by adopting more general non-linear and time varying functions. Smooth transition formulations for time-varying parameters have been used in the context of autoregressive models and are usually referred to as STAR models; see van Dijk et al. (2002) for an overview.

Our proposed class of trend-cycle models with smooth transition parameters is applied to four U.S. macroeconomic time series: gross domestic product, investment, unemployment and the industrial production index. We find that the parameters are not constant over the longer sample of 1948-2004 and that a better fit is obtained when smooth transition parameters are considered. In particular, the volatility of the business cycle associated with GDP decreases while the period of the cycle increases. However, this latter finding appears not to be strongly significant. The decrease of cycle volatility remains highly significant and is also observed in the other economic time series.

This chapter is organised as follows. In Section 4.2, the details of the trend-cycle decomposition model with its extension are discussed. Section 4.3 presents the state space representation of our model. Section 4.4 provides the empirical results for four U.S. macroeconomic time series. Conclusions are in Section 4.5.
4.2 Trend-cycle decomposition model

4.2.1 Fixed parameter specification

In this paper, we adopt the unobserved components (UC) approach to modelling macro-economic time series. The central decomposition in our investigations is given by

\[ y_t = \mu_t + \psi_t + \epsilon_t, \quad t = 1, \ldots, n, \]  

(4.1)

where \( y_t \) is the data, \( \mu_t \) represents a trend, \( \psi_t \) a cycle and \( \epsilon_t \) an idiosyncratic shock.

The trend is modelled as a higher order integrated random walk:

\[ \Delta^d \mu_t = \eta_t, \quad \eta_t \sim \text{NID}(0, \sigma^2_\eta) \]  

(4.2)

where \( d = 1, 2, \ldots \) A higher integration order \( d \) corresponds to a smoother trend. The trend is an \( I(d) \) process, and will be the main source of non-stationarity in the series. In practice, the most important cases are \( d = 1 \) (random walk) and \( d = 2 \) (smooth trend). When \( d > 2 \), the trend can be interpreted as a model representation of Butterworth type filters; see Gomez (2001). We model the shock \( \epsilon_t \) as a Gaussian white noise process:

\[ \epsilon_t \sim \text{NID}(0, \sigma^2_\epsilon). \]  

(4.3)

The disturbances \( \eta_t \) and \( \epsilon_t \) are mutually and serially independent at all lags.

The main focus of this chapter is the cycle component \( \psi_t \). Our basic approach follows Harvey (1989). The deterministic sine-cosine wave

\[ \psi_t = A \cos(\lambda t) + B \sin(\lambda t), \]  

(4.4)

is a cyclical function with frequency \( \lambda \) or period \( \omega = 2\pi/\lambda \), amplitude \( \sqrt{A^2 + B^2} \) and phase \( \tan^{-1}(B/A) \). It can be written in recursive form as

\[
\begin{pmatrix}
\psi_{t+1} \\
\dot{\psi}_{t+1}
\end{pmatrix} =
\begin{bmatrix}
\cos \lambda & \sin \lambda \\
-\sin \lambda & \cos \lambda
\end{bmatrix}
\begin{pmatrix}
\psi_t \\
\dot{\psi}_t
\end{pmatrix} +
\begin{pmatrix}
\kappa_t \\
\dot{\kappa}_t
\end{pmatrix},
\quad t = 1, \ldots, n,
\]  

(4.5)

with \( \psi_1 = A, \quad \dot{\psi}_1 = B \). It is rarely appropriate to model cycles in economic series with a deterministic trigonometric function. A stochastic cycle, based on this deterministic specification can be obtained by introducing a damping factor \( \phi \) and adding random shocks or disturbances in each step of the recursion:

\[
\begin{pmatrix}
\psi_{t+1} \\
\dot{\psi}_{t+1}
\end{pmatrix} = \phi
\begin{bmatrix}
\cos \lambda & \sin \lambda \\
-\sin \lambda & \cos \lambda
\end{bmatrix}
\begin{pmatrix}
\psi_t \\
\dot{\psi}_t
\end{pmatrix} +
\begin{pmatrix}
\kappa_t \\
\dot{\kappa}_t
\end{pmatrix},
\quad t = 1, \ldots, n,
\]  

(4.6)

\[
\begin{pmatrix}
\kappa_t \\
\dot{\kappa}_t
\end{pmatrix} \sim \text{NID}\left(0, \begin{bmatrix} \sigma^2_\kappa & 0 \\ 0 & \sigma^2_\kappa \end{bmatrix}\right),
\quad t = 1, \ldots, n.
\]  

(4.7)
The disturbance terms $\kappa_t$ and $\dot{\kappa}_t$ are mutually and serially independent of both $\eta_t$ and $\epsilon_t$. The stochastic specification contains the damping factor $\phi$, which is restricted to the range $[0, 1)$ to maintain stationarity. Values of $\phi$ close to one correspond to a more persistent cycle. It is easily derived that the cycle $\psi_t$ follows an ARMA(2,1) process

\[ \psi_t = \alpha_1 \psi_{t-1} + \alpha_2 \psi_{t-2} + \zeta_t + \theta \zeta_{t-1}, \]  

where $\alpha_1 = 2\phi \cos \lambda$, $\alpha_2 = -\phi^2$, $\theta = -\phi \cos \lambda$ and $\zeta_t$ is a Gaussian noise process

\[ \zeta_t \sim \text{NID} \left(0, \frac{\sigma_\kappa^2 (1 + \phi^2)}{1 + \phi^2 \cos^2 \lambda}\right). \]

The recursion (4.6) can be viewed as a method of specifying an ARMA process with restrictions on the coefficients such that the process exhibits cyclical behaviour. In particular, the autocorrelation of $\psi_t$ at l lags is given by

\[ \rho_l = \phi^l \cos(\lambda l) \]

and clearly contains a cyclical component. The unconditional variance of the cycle is given by

\[ \text{Var}(\psi_t) = \frac{\sigma_\kappa^2}{1 - \phi^2}. \]

### 4.2.2 Time-varying parameter specification

The cycle specification (4.6) contains three parameters that are assumed to be constant throughout the sample: the disturbance variance $\sigma_\kappa^2$, the period $\omega = 2\pi/\lambda$ and the damping factor $\phi$. We can generalise the model by allowing each of these parameters to change over time. As a general specification, we consider

\[
\sigma_{\kappa,t}^2 = f_\sigma(t), \\
\omega_t = f_\omega(t), \\
\phi_t = f_\phi(t),
\]

where $f_\theta(t)$ is some deterministic function of time for parameter $\theta$.

For the variance parameter, we consider the function

\[ f_\sigma(t) = \exp \left( c_\sigma + \gamma_\sigma \text{lt}(s_\sigma(t - \tau_\sigma)) \right), \]

\[ \text{lt}(x) = \frac{e^x}{1 + e^x}, \]

where the parameter $s_\theta$ determines the direction and acuteness of the change, while $\tau_\theta$ gives the mid-time position of the change of the function $f_\sigma(t)$. The argument of the
exponential function in (4.13) varies between $c_\sigma$ and $c_\sigma + e^{\gamma_\sigma}$. This specification can accommodate both gradual changes and structural breaks.

To distinguish the cycle from from a seasonal pattern, we impose the restriction that the period $\omega$ should be at least two. Thus, we use

$$f_\omega(t) = 2 + \exp\left(c_\omega + \gamma_\omega \ln(s_\omega(t - \tau_\omega))\right),$$

(4.15)

where $s_\omega$, $\tau_\omega$ and $\gamma_\omega$ have the same interpretation as before. The second term on the right hand side of equation (4.15) varies between $c_\omega$ and $c_\omega + e^{\gamma_\omega}$.

The time-varying parameter specifications for $\sigma_\kappa^2$, $\omega$ and $\phi$ require estimation of three extra parameters each. Since the position and the speed of changes (if any) are unknown parameters to be estimated from the data, it is considerably more flexible than the fixed parameter specification (4.6).

The logistic based smooth transition function is strictly monotonic. Since this imposes a restriction on the parameter evolution that is not necessarily supported by theory, we also consider a more flexible specification based on smooth spline functions. For the variance parameter, this specification is

$$f_\sigma(t) = \exp(c_\sigma + w_t \delta_\sigma),$$

(4.16)

where $c_\sigma$ is a constant, $\delta_\sigma$ is a column vector of coefficients and $w_t$ is a row vector of weights. The vector $w_t$ depends on the position of so-called knots and the distance between them. The precise construction of the weight vector can be found in Poirier (1973), Section 2. In this study we take five knots, which are equally distributed over time. Thus, two knots are placed at the first and the last data point and the remaining three knots are placed, equally distributed, between the sample endpoints. We restrict the first knot to be zero, such that $f_\sigma(t)$ varies between $e^{c_\sigma}$ and infinity. The spline specification for the period $\omega$ is given by:

$$f_\omega(t) = 2 + \exp(c_\omega + w_t \delta_\omega)$$

(4.17)

The spline specifications for the cycle variance or the cycle period require estimation of five extra parameters each (four coefficients and one constant). Note that $w_t$ is the same for all functions since the knot positions are fixed for all parameters. A different number of knots have also been considered but in the empirical applications we found that five knots to provide a good compromise between parsimony and fit.

In principle we can devise similar transformations for the damping factor $\phi_t$. However, as we shall see in the empirical section, the damping factor is generally quite stable over time. Therefore, we will restrict the discussion to $\sigma_\kappa^2$ and $\omega$. 

4.3 State space representation

As described in section 1.4 and 2.2, a standard univariate linear Gaussian state space model is defined by a measurement equation

\[ y_t = Z_t \alpha_t + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, \sigma^2_\epsilon) \]  

(4.18)

and a state equation

\[ \alpha_{t+1} = T_t \alpha_t + R \xi_t, \quad \xi_t \sim \text{NID}(0, Q_t). \]  

(4.19)

The state equation describes the evolution of the unobserved state vector \( \alpha_t \), which contains the relevant variables needed to model the dynamics of the process. The measurement equation relates the state vector to the actual observation. The matrices \( Z_t, T_t \) and \( Q_t \) are deterministic (possibly time-varying) matrices. In this formulation, the disturbance variance matrix \( Q_t \) is of full rank, even though the state will have many component without associated disturbance terms.

The unobserved component model that we employ in this chapter contains a higher order trend, a cycle component with time-varying parameters and observation noise. It is defined by equations (4.1), (4.2), (4.3) and (4.6), and can be readily written in state space form. The transition matrix is given by

\[ T_t = \begin{bmatrix} M & O \\ O & C_t \end{bmatrix}, \]  

(4.20)

\[ M = I_d + \begin{bmatrix} O & I_{d-1} \\ 0 & O \end{bmatrix}, \quad C_t = \phi \begin{bmatrix} \cos \lambda_t & \sin \lambda_t \\ -\sin \lambda_t & \cos \lambda_t \end{bmatrix}. \]  

(4.21)

The state disturbance variance matrix \( Q_t \) and selection matrix \( R \) are

\[ Q_t = \begin{bmatrix} \sigma^2_\eta & 0 & 0 \\ 0 & \sigma^2_{\kappa,t} & 0 \\ 0 & 0 & \sigma^2_{\kappa,t} \end{bmatrix}, \quad R = \begin{bmatrix} O \\ I_3 \end{bmatrix}. \]  

(4.22)

The measurement equation is time-invariant and is given by

\[ y_t = Z \alpha_t + \epsilon_t, \]  

(4.23)

\[ Z = \begin{bmatrix} 1 & O & 1 & 0 \end{bmatrix}. \]  

(4.24)

In the partitioned matrices, \( I_d \) represents a \( d \)-dimensional identity matrix, while \( O \) is a conforming zero-matrix. The state \( \alpha_t \) is the \( d + 2 \) dimensional vector

\[ \alpha_t = \begin{bmatrix} \mu_t & \Delta \mu_t & \ldots & \Delta^{d-1} \mu_t & \psi_t & \dot{\psi}_t \end{bmatrix}. \]  

(4.25)
CHAPTER 4. CYCLES WITH SMOOTH-TRANSITION PARAMETERS

The time variation in the parameters $\sigma_{\kappa,t}^2$ and $\omega_t = 2\pi/\lambda_t$ were described in section 4.2.2.

Although the parameters in this model vary as non-linear functions of time, this model is from a filtering point of view a linear state space model, as the non-linearity only acts on a deterministic part of the model. Estimation of the mean and variance of the state in this model can therefore be done via the standard (diffuse) Kalman filter and Kalman smoother of section 2.2. The to accommodate the slightly different formulation of the state disturbance term, the variance matrix $H_t$ in the filtering equations are replaced by $RQ_t R'$. The extra parameters which characterise the smooth transition functions section 4.2.2 are estimated using maximum likelihood via the standard prediction error decomposition (2.18).

4.4 Empirical evidence from U.S. economic time series

4.4.1 Data

In the empirical study we consider four key time series from the U.S. economy: gross domestic production (GDP), fixed private investment (FPI), unemployment (UNRATE) and the industrial production index (INDPRO). The sample ranges from 1948 to 2004, first half-year. The series GDP and FPI are sampled at a quarterly frequency and compiled by the Bureau of Economic Analysis, while INDPRO and UNRATE are monthly data originating from the Federal Reserve Bank and Bureau of Labor Statistics respectively. The series can be obtained from the publicly accessible FRED database of the Federal Reserve Bank of St. Louis at http://research.stlouisfed.org/fred2. All series were seasonally adjusted by the data originators. Table 4.4.1 provides further details of the source and definitions of our dataset.

<table>
<thead>
<tr>
<th>series</th>
<th>frequency</th>
<th>data range</th>
<th>description</th>
</tr>
</thead>
</table>

Table 4.1: Data description.
### 4.4. Empirical Evidence from U.S. Economic Time Series


<table>
<thead>
<tr>
<th></th>
<th>$\sigma_\epsilon$</th>
<th>$\sigma_\mu$</th>
<th>$\sigma_\kappa$</th>
<th>$\lambda$</th>
<th>$\phi$</th>
<th>$\sigma_\lambda$</th>
<th>$\sigma_\phi$</th>
<th>$\mu_\phi$</th>
<th>$\lambda_\phi$</th>
<th>$\phi_\phi$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\gamma$</th>
<th>$\delta$</th>
<th>$\epsilon$</th>
</tr>
</thead>
<tbody>
<tr>
<td>full</td>
<td>$2.11 \times 10^{-6}$</td>
<td>0.001</td>
<td>0.017</td>
<td>0.359</td>
<td>0.900</td>
<td>$17.82$</td>
<td>$20.86$</td>
<td>$719.37$</td>
<td>$-1428.46$</td>
<td>$-1411.63$</td>
<td>517.10</td>
<td>-1023.92</td>
<td>-1007.09</td>
<td>57.93</td>
<td>-146.86</td>
</tr>
<tr>
<td>GDP</td>
<td>I</td>
<td>$9.71 \times 10^{-8}$</td>
<td>6.97 $\times 10^{-8}$</td>
<td>0.033</td>
<td>0.144</td>
<td>0.945</td>
<td>$1.10$</td>
<td>$33.55$</td>
<td>$331.55$</td>
<td>$-652.52$</td>
<td>$-639.50$</td>
<td>243.37</td>
<td>476.18</td>
<td>463.15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>$2.56 \times 10^{-8}$</td>
<td>1.84 $\times 10^{-8}$</td>
<td>0.021</td>
<td>0.177</td>
<td>0.951</td>
<td>$22.30$</td>
<td>$28.03$</td>
<td>$390.32$</td>
<td>$-770.09$</td>
<td>$-756.96$</td>
<td>279.92</td>
<td>549.29</td>
<td>536.17</td>
<td></td>
</tr>
<tr>
<td></td>
<td>full</td>
<td>$1.17 \times 10^{-6}$</td>
<td>0.003</td>
<td>0.054</td>
<td>0.312</td>
<td>0.938</td>
<td>$15.63$</td>
<td>$50.61$</td>
<td>$517.10$</td>
<td>$-1023.92$</td>
<td>$-1007.09$</td>
<td>97.66</td>
<td>185.23</td>
<td>162.72</td>
<td></td>
</tr>
<tr>
<td>FPI</td>
<td>I</td>
<td>$2.76 \times 10^{-7}$</td>
<td>0.004</td>
<td>0.047</td>
<td>0.476</td>
<td>0.921</td>
<td>$8.46$</td>
<td>$24.49$</td>
<td>$243.37$</td>
<td>$-476.18$</td>
<td>$-463.15$</td>
<td>-1.15</td>
<td>12.48</td>
<td>31.39</td>
<td></td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>$9.87 \times 10^{-7}$</td>
<td>0.004</td>
<td>0.045</td>
<td>0.285</td>
<td>0.947</td>
<td>$7.17$</td>
<td>$29.60$</td>
<td>$279.92$</td>
<td>$-549.29$</td>
<td>$-536.17$</td>
<td>130.65</td>
<td>251.12</td>
<td>232.12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>full</td>
<td>0.070</td>
<td>0.013</td>
<td>0.741</td>
<td>0.131</td>
<td>0.976</td>
<td>$67.18$</td>
<td>$47.57$</td>
<td>$97.66$</td>
<td>$-185.23$</td>
<td>$-162.72$</td>
<td>1015.57</td>
<td>2020.96</td>
<td>2002.06</td>
<td></td>
</tr>
<tr>
<td>URATE</td>
<td>I</td>
<td>0.080</td>
<td>0.007</td>
<td>0.941</td>
<td>0.144</td>
<td>0.980</td>
<td>$27.30$</td>
<td>$29.13$</td>
<td>-1.15</td>
<td>12.48</td>
<td>31.39</td>
<td>63.09</td>
<td>$1015.57$</td>
<td>2020.96</td>
<td>2002.06</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>0.071</td>
<td>0.017</td>
<td>0.461</td>
<td>0.120</td>
<td>0.971</td>
<td>$16.06$</td>
<td>$19.54$</td>
<td>$130.65$</td>
<td>$-251.12$</td>
<td>$-232.12$</td>
<td>1258.02</td>
<td>2505.86</td>
<td>2486.87</td>
<td></td>
</tr>
<tr>
<td></td>
<td>full</td>
<td>$1.66 \times 10^{-6}$</td>
<td>$7.93 \times 10^{-4}$</td>
<td>0.030</td>
<td>0.175</td>
<td>0.964</td>
<td>$322.16$</td>
<td>$93.02$</td>
<td>$2206.89$</td>
<td>$-4403.69$</td>
<td>$-4381.18$</td>
<td>1015.57</td>
<td>2020.96</td>
<td>2002.06</td>
<td></td>
</tr>
<tr>
<td>INPRO</td>
<td>I</td>
<td>$2.14 \times 10^{-7}$</td>
<td>$4.46 \times 10^{-4}$</td>
<td>0.044</td>
<td>0.167</td>
<td>0.973</td>
<td>$109.84$</td>
<td>$63.09$</td>
<td>$1015.57$</td>
<td>$-2020.96$</td>
<td>$-2002.06$</td>
<td>1258.02</td>
<td>2505.86</td>
<td>2486.87</td>
<td></td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>$5.30 \times 10^{-5}$</td>
<td>$7.39 \times 10^{-4}$</td>
<td>0.016</td>
<td>0.164</td>
<td>0.946</td>
<td>$37.52$</td>
<td>$17.88$</td>
<td>$1258.02$</td>
<td>$-2505.86$</td>
<td>$-2486.87$</td>
<td>1258.02</td>
<td>2505.86</td>
<td>2486.87</td>
<td></td>
</tr>
</tbody>
</table>
4.4.2 Basic decompositions

The estimation results for the basic model without time-varying parameters are presented in Table 4.4.1 under the header ‘full’. The trend components are specified as order two integrated random walks. The associated variances for the trends in the four economic time series are estimated as small values resulting in smooth trend functions. The relatively smallest trend variance is obtained for the GDP series. Further, the irregular variances are very small for the GDP, INDPRO and the FPI series. Such results are typical for aggregated macroeconomic series of stable and large economies. Most of the variation in the series is due to the cycle component as this variance is largest in all cases. The periods of the cycles are estimated as approximately 8 years (GDP), 5 years (FPI, UNRATE) and 3 years (INDPRO). The persistence of the cycles is determined by the damping factor \( \phi \) and its estimates range from 0.90 to 0.94 on a quarterly basis\(^1\). Such values are satisfactory since an adequate level of persistence in the cycle leads to an autocorrelation function that exhibits a cyclical pattern and a forecast function that contains sufficient cyclical variations.

The main aim of this chapter is to study the stability of parameter values when the model is applied to a sample that covers a long time span. In order to obtain a first indication, we re-estimate the decomposition model for the four economic series based on two subsamples that consist of observations from the first and second halves of the full sample. The estimation results for the two subsamples are presented in Table 4.4.1 under the headers ‘I’ and ‘II’. It is revealing that the estimated parameter coefficients are in most cases very different for the two subsamples. Our focus is on the cycle component and therefore particular attention is given to the coefficients \( \sigma_\kappa \), \( \lambda \) and \( \phi \). The variance and the frequency of the cycle are estimated quite differently in the two samples while persistence has remained approximately constant. The implied periods of the cycles of the FPI, UNRATE and INPRO series have increased in the second sample leading to longer cycles. For the GDP series the estimated period of the cycle has become shorter. Although in the period after the 1970s the cycle in GDP has become longer (and weaker), the swing of the cycle in the 1950s and the early 1960s was relatively long and persistent. The typical GDP cycle length of five to seven years has only been observed in the late 1960s and 1970s. In all cases however, it is found that the variation of the cycle in the early sample is larger compared to the cycle variation in the second sample.

Diagnostic checking for our model can be performed by testing the standardised prediction errors

\[
v_t / \sqrt{F_t}
\]

\(^1\)In the case of a monthly series, the autoregressive coefficient on a quarterly basis is taken as \( \phi^3 \).
obtained from the Kalman filter of section 2.2. The Normality statistic $N$ that we use is the asymptotic test by Bowman and Shenton (1975), based on the sample skewness and kurtosis, which is also known as the Jarque-Bera test. The heteroskedasticity statistic $H$ is the Goldfeld-Quandt test, which is calculated as the ratio between the sum of squared prediction errors in two exclusive subsets of the sample. Finally the serial correlation statistic $Q(l)$ is the standard portmanteau test based on the sum of $l$ squared sample autocorrelations derived by Ljung and Box (1978). The $p$-values of the tests are given between brackets in the table.

Most of the diagnostic statistics in Table 4.4.1 are rather significant. Splitting the data in two subsamples improves the statistics in most cases. We attempt to use a common model for each of the series in this chapter. As a result, for some series we shall not eliminate the significance of all the test statistics.

### 4.4.3 Smooth transitions over time

We first analyse the GDP (quarterly) and unemployment (monthly) series in more detail, using the trend-cycle decomposition model with time-varying period and disturbance variance. The logistic and spline transition functions that we use were given in equations (4.13)–(4.17). The coefficients of the smooth transition functions are estimated jointly with the other parameters of the model by the direct maximisation of the likelihood function with respect to the parameter vector. For the spline specification, five equidistant knots were chosen and estimated by maximum likelihood. The resulting parameter estimates are reported in Tables 4.3 and 4.4, with standard errors in parentheses. Figures 4.1 and 4.2 show the estimated cycles and the evolutions of the cycle variance and the period as functions of time. The order of the trend component is equal to two for the smooth transition specification and three for the spline specification. These smoothness orders are chosen based on the fit of the data. We should note that higher integration order $d$ produces a smoother trend but permits more flexibility in the other components. Since the full specification has a large number of parameters, we only provide the full estimation results for quarterly GDP and monthly unemployment. Similar results have been obtained for the other two series.

First, we notice a substantial increase in the log-likelihood values for both series, compared the the time-invariant specification in the previous subsection. The GDP likelihood increases with 19 and 22 points, while the unemployment likelihoods improve with 49 and 55 points, for the spline and smooth transition specification respectively. We use the corrected Akaike Information Criterion (AICC), and the Bayes Information Criterion (BIC) to evaluate model fit; details can be found in Akaike (1974), Schwarz (1978) and Hurvich and Tsai (1991). Both the AICC and BIC values from the unem-
<table>
<thead>
<tr>
<th></th>
<th>Spline specification</th>
<th>Logit specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>irregular var.</td>
<td>$\sigma_\epsilon$</td>
<td>0.00096</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0013)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.01 × 10^{-5})</td>
</tr>
<tr>
<td>trend var.</td>
<td>$\sigma_\eta$</td>
<td>0.00013</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.2015)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.01 × 10^{-5})</td>
</tr>
<tr>
<td>cycle var. coeff.</td>
<td>$c_\sigma$</td>
<td>0.00021</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.66 × 10^{-9})</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.2015)</td>
</tr>
<tr>
<td></td>
<td>$\delta_{\sigma 2}$</td>
<td>-1.755</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.2044)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0025)</td>
</tr>
<tr>
<td></td>
<td>$\delta_{\sigma 3}$</td>
<td>-0.726</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.3538)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0002)</td>
</tr>
<tr>
<td></td>
<td>$\delta_{\sigma 4}$</td>
<td>-2.840</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.6872)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.46 × 10^{-7})</td>
</tr>
<tr>
<td></td>
<td>$\delta_{\sigma 5}$</td>
<td>-2.025</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.6872)</td>
</tr>
<tr>
<td>period coeff.</td>
<td>$\gamma_\omega$</td>
<td>15.604</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(7.2446)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.2015)</td>
</tr>
<tr>
<td></td>
<td>$\delta_{\omega 2}$</td>
<td>-0.481</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.5705)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0256)</td>
</tr>
<tr>
<td></td>
<td>$\delta_{\omega 3}$</td>
<td>0.490</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.5350)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0459)</td>
</tr>
<tr>
<td></td>
<td>$\delta_{\omega 4}$</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.5505)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.3869)</td>
</tr>
<tr>
<td></td>
<td>$\delta_{\omega 5}$</td>
<td>0.528</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.4065)</td>
</tr>
<tr>
<td>damping factor</td>
<td>$\phi$</td>
<td>0.914</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0260)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0244)</td>
</tr>
<tr>
<td>diagnostics</td>
<td>$N$</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.762]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.003]</td>
</tr>
<tr>
<td></td>
<td>$H$</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.844]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.379]</td>
</tr>
<tr>
<td></td>
<td>$Q(20)$</td>
<td>18.60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.548]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.411]</td>
</tr>
<tr>
<td>likelihood</td>
<td>Lik</td>
<td>738.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AICC</td>
<td>-1446.72</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BIC</td>
<td>-1400.82</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.3: Estimation results for the GDP series. The volatility and the period of the cycle are time-varying. For the spline specification, the integration order of the trend is equal to three. For the logistic specification we have $d = 2$. 
## 4.4. Empirical Evidence from U.S. Economic Time Series

Spline specification  | Logit specification
--- | ---
irregular var.  |  |  
| \( \sigma_\epsilon \) | 0.0013 | \( \sigma_\epsilon \) | 0.061 \\
|  | (0.4154) |  | (0.0080)

| trend var.  |  |  
| \( \sigma_\eta \) | \( 4.32 \times 10^{-5} \) (1.62\times10^{-5}) | \( \sigma_\eta \) | 0.011 \\
|  | \( (0.1106) \) |  | \( (0.0058) \)

cycle var. coeff.  |  |  
| \( c_\sigma \) | 0.182 | \( c_\sigma \) | 1.430 \\
|  | (0.0057) |  | (0.2934)

|  |  |  
| \( \delta_{\sigma_2} \) | \(-1.602 \) | \( \tau_\sigma \) | 0.0092 \\
|  | (0.1128) |  | \( (0.0025) \)

|  |  |  
| \( \delta_{\sigma_3} \) | \(-1.687 \) | \( s_\sigma \) | 0.0092 \\
|  | (0.1228) |  | \( (0.0025) \)

|  |  |  
| \( \delta_{\sigma_4} \) | \(-2.066 \) | \( \gamma_\sigma \) | \(-9.042 \) \\
|  | (0.1350) |  | \( (0.8295) \)

|  |  |  
| \( \delta_{\sigma_5} \) | \(-2.654 \) |  |  \\
|  | (0.3009) |  | (0.0019)

| period coeff.  |  |  
| \( \gamma_\omega \) | \( 40.961 \) (15.149) | \( c_\omega \) | 4.253 \\
|  | (0.5375) |  | (0.7534)

|  |  |  
| \( \delta_{\omega_2} \) | 1.275 | \( \tau_\omega \) | 0.0002 \\
|  | (0.3873) |  | (0.0025)

|  |  |  
| \( \delta_{\omega_3} \) | 0.383 | \( s_\omega \) | 0.0002 \\
|  | (0.4649) |  | (0.0025)

|  |  |  
| \( \delta_{\omega_4} \) | 1.235 | \( \gamma_\omega \) | \(-0.591 \) \\
|  | (0.4649) |  | \( (1.4189) \)

|  |  |  
| \( \delta_{\omega_5} \) | 0.592 |  |  \\
|  | (0.5464) |  | (0.0019)

damping factor  |  |  
| \( \phi \) | 0.985 | \( \phi \) | 0.976 \\
|  | (0.0048) |  | (0.0055)

| diagnostics  |  |  
| \( N \) | 28.76 | \( N \) | 95.62 \\
|  | (0.0000) |  | (0.0000)

|  |  |  
| \( H \) | 0.89 | \( H \) | 1.13 \\
|  | (0.404) |  | (0.347)

|  |  |  
| \( Q(20) \) | 71.45 | \( Q(20) \) | 34.72 \\
|  | (0.0000) |  | (0.022)

| likelihood  |  |  
| \( \text{Lik} \) | 147.00 | \( \text{Lik} \) | 152.24 \\
|  | \( \text{AICC} \) | -265.37 | \( \text{AICC} \) | -282.08 \\
|  | \( \text{BIC} \) | -202.73 | \( \text{BIC} \) | -232.77 \\

Table 4.4: Estimation results for the unemployment series. The volatility and the period of the cycle are time-varying. For the spline specification, the integration order of the trend is equal to three. For the logistic specification we have \( d = 2 \).

Figure 4.1: Estimated cycle $\psi_t$, time-varying cycle volatility $f_\sigma(t)$ and time-varying period $f_\omega(t)$ of the GDP series are respectively given in the first, second and third row.
4.4. EMPIRICAL EVIDENCE FROM U.S. ECONOMIC TIME SERIES

Figure 4.2: Estimated cycle $\psi_t$, time-varying cycle volatility $f_\sigma(t)$, and time-varying period $f_\omega(t)$ of the unemployment series are respectively given in the first, second and third row.
ployment series suggest that it is preferable to allow the cycle parameters $\sigma$, $\kappa$, and $\omega$ to vary over time. The conclusion from the GDP series is ambiguous: we would prefer a fixed parameter specification based on the BIC, but a time varying specification based on the AICC.

Second, from the figures it can be seen that the variation in the cycle variance is quite large: for both series the cycle variation at the end of the sample is reduced to less than half the value at the beginning of the sample. This can be seen in the plots from both the spline and smooth transition specifications.

Third, the period in both series shows a generally rising trend, varying roughly between four and twelve years. The GDP cycles for the logistic and spline specification are similar. However, for the unemployment series these cycles are clearly different. Due to the asymmetric features of the unemployment series, it is evidently difficult to model this time series within the current framework. With the logistic specification, we obtain a cycle with the period varying between three and six years. The period is small compared to the period of the cycle obtained from the model with the spline specification. The spline specification shows a smooth wave-like pattern for the period, varying between four and twelve years. The diagnostic tests indicates that the spline specification gives a better fit for the GDP series than the logistic specification. The results for unemployment are ambiguous: the spline specification is better in terms of Normality but worse with respect to the serial correlation.

For the GDP series, the time-varying specifications result in residuals which are closer to Normality and heteroskedasticity. There is little improvement in serial correlation, but even in the original specification the Q statistic was not significant. The unemployment residuals show improvement in serial correlation and heteroskedasticity, but Normality is still rejected. For this series there remains a significant amount of serial correlation in the residuals in both the fixed parameter and the time-varying specification.

As a final specification and robustness check, we estimate the trend-cycle model with time-varying cycle disturbance variance, but with a fixed period parameter. The estimation results from all four series are given in Table 4.5 and 4.6. The integration order of the trend is two for both time-varying specifications and all series. The AICC and BIC indicate that for all series both time-varying specifications are preferred to the fixed parameter model. Except for the autocorrelation in the GDP residuals, all diagnostic statistics of the time-varying models improve on their fixed parameter counterparts. In most cases, the spline specifications give slightly better results than the smooth transition functions. Figures 4.4 and 4.3 show the evolution of the cycle disturbance variance. It is clear from the plots that the qualitative conclusions drawn from both specifications are similar.
4.4. EMPIRICAL EVIDENCE FROM U.S. ECONOMIC TIME SERIES

<table>
<thead>
<tr>
<th></th>
<th>GDP</th>
<th>IN</th>
<th>U</th>
<th>IPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma_\epsilon )</td>
<td>2.70 \times 10^{-6} \hspace{1cm} (2.70 \times 10^{-12})</td>
<td>1.08 \times 10^{-6} \hspace{1cm} (2.14 \times 10^{-14})</td>
<td>0.061 \hspace{1cm} (0.0083)</td>
<td>4.04 \times 10^{-5} \hspace{1cm} (3.48 \times 10^{-9})</td>
</tr>
<tr>
<td>( \sigma_\eta )</td>
<td>0.0012 \hspace{1cm} (2.30 \times 10^{-4})</td>
<td>0.003 \hspace{1cm} (6.63 \times 10^{-4})</td>
<td>0.012 \hspace{1cm} (0.0019)</td>
<td>6.60 \times 10^{-4} \hspace{1cm} (1.06 \times 10^{-4})</td>
</tr>
<tr>
<td>( c_\sigma )</td>
<td>0.00022 \hspace{1cm} (1.87 \times 10^{-9})</td>
<td>0.0017 \hspace{1cm} (2.10 \times 10^{-7})</td>
<td>0.184 \hspace{1cm} (0.0058)</td>
<td>2.40 \times 10^{-4} \hspace{1cm} (1.39 \times 10^{-8})</td>
</tr>
<tr>
<td>( \delta_{\sigma_2} )</td>
<td>-1.427 \hspace{1cm} (0.2300)</td>
<td>-1.843 \hspace{1cm} (0.2127)</td>
<td>-2.094 \hspace{1cm} (0.1269)</td>
<td>-0.973 \hspace{1cm} (0.0555)</td>
</tr>
<tr>
<td>( \delta_{\sigma_3} )</td>
<td>-0.913 \hspace{1cm} (0.1812)</td>
<td>-1.287 \hspace{1cm} (0.1848)</td>
<td>-1.931 \hspace{1cm} (0.1072)</td>
<td>-1.560 \hspace{1cm} (0.1171)</td>
</tr>
<tr>
<td>( \delta_{\sigma_4} )</td>
<td>-2.779 \hspace{1cm} (0.3098)</td>
<td>-2.601 \hspace{1cm} (0.2784)</td>
<td>-2.807 \hspace{1cm} (0.1492)</td>
<td>-2.405 \hspace{1cm} (0.1171)</td>
</tr>
<tr>
<td>( \delta_{\sigma_5} )</td>
<td>-2.416 \hspace{1cm} (0.4419)</td>
<td>-2.961 \hspace{1cm} (0.5623)</td>
<td>-3.357 \hspace{1cm} (0.3024)</td>
<td>-2.858 \hspace{1cm} (0.2139)</td>
</tr>
</tbody>
</table>

\( \lambda \) \hspace{1cm} 0.328 \hspace{1cm} (0.0027) \hspace{1cm} 0.299 \hspace{1cm} (0.0013) \hspace{1cm} 0.122 \hspace{1cm} (0.0002) \hspace{1cm} 0.150 \hspace{1cm} (0.0002)

\( \omega \) (in years) \hspace{1cm} 4.85 \hspace{1cm} 5.26 \hspace{1cm} 4.31 \hspace{1cm} 3.48

\( \phi \) \hspace{1cm} 0.895 \hspace{1cm} (0.0270) \hspace{1cm} 0.946 \hspace{1cm} (0.0167) \hspace{1cm} 0.974 \hspace{1cm} (0.0060) \hspace{1cm} 0.963 \hspace{1cm} (0.0086)

<table>
<thead>
<tr>
<th>N</th>
<th>1.44 \hspace{1cm} [0.486]</th>
<th>1.11 \hspace{1cm} [0.574]</th>
<th>24.39 \hspace{1cm} [0.000]</th>
<th>130.35 \hspace{1cm} [0.000]</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>0.89 \hspace{1cm} [0.646]</td>
<td>0.95 \hspace{1cm} [0.835]</td>
<td>1.04 \hspace{1cm} [0.757]</td>
<td>0.94 \hspace{1cm} [0.642]</td>
</tr>
<tr>
<td>Q(20)</td>
<td>25.21 \hspace{1cm} [0.194]</td>
<td>28.01 \hspace{1cm} [0.109]</td>
<td>40.06 \hspace{1cm} [0.005]</td>
<td>50.15 \hspace{1cm} [0.000]</td>
</tr>
</tbody>
</table>

Lik 747.27 542.95 161.16 2307.17
AICC -1475.70 -1067.06 -304.06 -4596.08
BIC -1445.75 -1037.11 -263.66 -4555.67

Table 4.5: Estimation results: decomposition model using the spline specification to vary \( f_\sigma(t) \) over time. The integration order of the trend is equal to two.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>GDP</th>
<th>IN</th>
<th>U</th>
<th>IPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma_\epsilon )</td>
<td>( 1.23 \times 10^{-5} ) ((5.95 \times 10^{-11}))</td>
<td>( 4.91 \times 10^{-6} ) ((6.26 \times 10^{-13}))</td>
<td>( 0.063 ) ((0.0078))</td>
<td>( 8.57 \times 10^{-4} ) ((6.19 \times 10^{-4}))</td>
</tr>
<tr>
<td>( \sigma_\eta )</td>
<td>( 0.001 ) ((2.41 \times 10^{-4}))</td>
<td>( 0.003 ) ((6.67 \times 10^{-4}))</td>
<td>( 0.012 ) ((0.0020))</td>
<td>( 7.36 \times 10^{-4} ) ((1.11 \times 10^{-4}))</td>
</tr>
<tr>
<td>( c_\sigma )</td>
<td>( 0.021 ) ((1.63 \times 10^{-7}))</td>
<td>( 0.056 ) ((6.15 \times 10^{-7}))</td>
<td>( 1.430 ) ((0.2932))</td>
<td>( 0.569 ) ((2.22 \times 10^{-5}))</td>
</tr>
<tr>
<td>( \tau_\sigma )</td>
<td>( 0.067 ) ((0.0038))</td>
<td>( 0.035 ) ((0.0035))</td>
<td>( 0.005 ) ((0.0061))</td>
<td>(-0.013 ) ((0.0019))</td>
</tr>
<tr>
<td>( s_\sigma )</td>
<td>(-0.011 ) ((0.0021))</td>
<td>(-0.009 ) ((0.0017))</td>
<td>(-0.002 ) ((0.0002))</td>
<td>(-0.002 ) ((0.0001))</td>
</tr>
<tr>
<td>( \gamma_\sigma )</td>
<td>(-4.305 ) ((0.1726))</td>
<td>(-4.696 ) ((0.1756))</td>
<td>(-9.102 ) ((0.8271))</td>
<td>(-8.541 ) ((0.1142))</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>( 0.296 ) ((0.0019))</td>
<td>( 0.287 ) ((0.0011))</td>
<td>( 0.120 ) ((0.0002))</td>
<td>( 0.150 ) ((0.0002))</td>
</tr>
<tr>
<td>( \omega ) (in years)</td>
<td>( 5.30 )</td>
<td>( 5.47 )</td>
<td>( 4.35 )</td>
<td>( 3.48 )</td>
</tr>
<tr>
<td>( \phi )</td>
<td>( 0.908 ) ((0.0246))</td>
<td>( 0.952 ) ((0.0133))</td>
<td>( 0.974 ) ((0.0058))</td>
<td>( 0.963 ) ((0.0086))</td>
</tr>
<tr>
<td>( N )</td>
<td>( 9.40 ) ([0.009])</td>
<td>( 0.65 ) ([0.724])</td>
<td>( 28.22 ) ([0.000])</td>
<td>( 108.03 ) ([0.000])</td>
</tr>
<tr>
<td>( H )</td>
<td>( 0.86 ) ([0.503])</td>
<td>( 0.98 ) ([0.931])</td>
<td>( 1.37 ) ([0.019])</td>
<td>( 0.98 ) ([0.887])</td>
</tr>
<tr>
<td>( Q(20) )</td>
<td>( 24.45 ) ([0.223])</td>
<td>( 37.88 ) ([0.009])</td>
<td>( 36.45 ) ([0.014])</td>
<td>( 46.78 ) ([0.000])</td>
</tr>
<tr>
<td>( \text{Lik} )</td>
<td>( 738.74 )</td>
<td>( 535.79 )</td>
<td>( 150.55 )</td>
<td>( 2307.29 )</td>
</tr>
<tr>
<td>( \text{AICC} )</td>
<td>(-1460.81 )</td>
<td>(-1054.92 )</td>
<td>(-384.88 )</td>
<td>(-4598.37 )</td>
</tr>
<tr>
<td>( \text{BIC} )</td>
<td>(-1434.11 )</td>
<td>(-1028.22 )</td>
<td>(-248.95 )</td>
<td>(-4562.43 )</td>
</tr>
</tbody>
</table>

Table 4.6: Estimation results: decomposition model using the logistic specification to vary \( f_\sigma(t) \) over time. The integration order of the trend is equal to two.
Figure 4.3: Estimated cycle $\psi_t$ and time-varying cycle volatility $f_{\sigma}(t)$ using the spline specification. The results for GDP, Investment, Unemployment and IPI are respectively given in the first, second, third and fourth row.
In Figure 4.4: Estimated cycle $\psi_t$ and time-varying cycle volatility $f_t(\sigma)$ using the logistic specification. The results for GDP, Investment, Unemployment and IPI are respectively given in the first, second, third and fourth row.
4.5 Discussion and conclusion

In this chapter we have investigated the parameter stability of unobserved components time series models for a set of relatively long macroeconomic time series from 1948 to 2004. For this purpose, we considered flexible non-linear functions for parameter changes. For example, we can consider the model to have different parameters before time $\tau$ and after time $\tau$. Since such abrupt changes in parameters that determine the dynamics of economic time series are not likely, we prefer smooth transitions to different parameter specifications. These extensions for the basic trend-cycle decompositions are implemented and used for the modelling of four U.S. macro-economic time series: GDP, investment, unemployment and industrial production. It is shown that the fit of the smooth transition unobserved components time series model for these four time series has increased significantly compared to the basic model. The main conclusion is that the volatility of the cycle component has considerably reduced over 1948-2004. We also found some evidence that the cycle period has not remained constant over time.
Chapter 5

Asymmetric Business Cycles

Abstract

In macro-economic analysis, a standard decomposition considers times series as a combination of trend, cycle, season and irregular components. The class of unobserved components state space models provide statistical tools and theory to calculate such a decomposition. In this chapter we consider a non-linear unobserved components model which allows the cycle component to take an asymmetric shape. In theoretical and empirical studies, the asymmetry of cyclical behavior is often discussed and considered for series such as unemployment and gross domestic product. This chapter develops a limited modification of the standard cycle component which leads to a flexible device for asymmetric cycles. The presence of asymmetry can be tested using classical likelihood based test statistics. It is found in the empirical investigation that cyclical asymmetry is a prominent feature in the U.S. economy.

5.1 Introduction

Many aggregate economic time series exhibit cyclical fluctuations. Filters derived from a frequency domain representation of the series are often used in cycle analysis. The popular Hodrick and Prescott (1997) filter, despite its somewhat arbitrary nature, remains widely used, alongside refinements and improvements such as the Baxter and King (1999) and Christiano and Fitzgerald (2003) filters. These band-pass filters are usually designed to isolate the fluctuating components in the series with periods between six and thirty-two quarters. Slower moving components are classified as trend, while faster fluctuations comprise the irregular and seasonal parts of the series. The extracted cycles from band-pass filters are visually appealing, but their optimality characteristics typically break down near the end-points of the series. Most applications of these
filters are found in historic cycle analysis, although some constructions for forecasting
and confidence bounds have been proposed, e.g. Johnson and Gallego (2003).

In traditional linear autoregressive integrated moving average (ARIMA) models,
cyclical behaviour is usually implied by estimated model parameters rather than ex-
licitly modelled. The theory on estimation, testing, forecasting and building confidence
intervals is well established in ARIMA modelling, but a decomposition of the trend and
cycle is not made as explicit as in the frequency domain. In ARIMA modelling, the
trend is usually eliminated by differencing, resulting in models on growth variables.
Cyclical variation in the growth can be inferred from the serial correlation structure.
From a frequency domain point of view, taking first differences can be regarded as a
low-pass filter which does not separate the cyclical variation from the higher frequency
components.

Structural time series or unobserved components (UC) models represent an attrac-
tive alternative time domain modelling technique. Trend, cycles and higher frequency
components are explicitly modelled by stochastic processes and estimated from the data
using Kalman filter and smoothing algorithms. Similar to filters in the frequency do-
main, decompositions of separate components are immediately visible, while rigorous
methods for estimation, testing and forecasting are well developed. The common cycle
specification for macro-economic time series in UC models is constructed from stochas-
tic trigonometric functions, as described by Harvey and Jaeger (1993). A generalisation
of this specification was studied by Harvey and Trimbur (2003). A higher integration
order of the cycle was shown to result in a better approximation to the ideal band-pass
filter. In this chapter, we extend the basic stochastic trigonometric cycle specification
to account for asymmetries in the cycle.

It is widely believed that the cycle in many economic series are asymmetric in the
sense that the expansions and contractions do not occur with the same speed. An early
widely cited quantitative study on the asymmetry in economic cycles was published
by Neftci (1984). Given a time series \( y_t \), a Markov process \( I_t \) is defined with states
representing increases and decreases in \( y_t \). Neftci derives likelihood-based asymmetry
tests and posterior odds ratios from the transition probabilities of \( I_t \), but the series \( y_t \) is
not explicitly modelled. In the empirical investigation significant evidence of asymmetry
is found in unemployment rate series of the U.S.

Most of the subsequent work on asymmetric cycles concentrate on the U.S. gross
national product (GNP) series. Hamilton (1989) analysed the post-war U.S. GNP
series in an influential article using a non-linear parametric model, specifically, an
ARIMA\((r, 1, 0)\) model augmented with an latent Markov switching trend process. The
chapter mainly focuses on filtering the unobserved regime from the data, and presents
evidence of the superiority of the specification compared to linear ARIMA and UC mod-

5.2 FORMULATION OF ASYMMETRIC STOCHASTIC CYCLE

Although Hamilton presents his model as an extension of Neftci’s approach, the issue of asymmetry is hinted at but not addressed explicitly. More recently, Clements and Krolzig (2003) developed rigorous tests for asymmetry in the Markov switching framework, analysing GNP, investment and consumption growth data from the U.S.


State space models with asymmetric cycles have been employed by Kim and Nelson (1999), Luginbuhl and de Vos (1999), Jesus Crespo Cuaresma (2004), either in combination with Markov switching, or using two regimes based on constructed variables or deterministic functions of past observations. Another example of state space modelling with two cycle regimes is given by Harvey (1989, section 6.5), who based the cycle frequency on the direction of the filtered cycle. Acemoglu and Scott (1997) constructed a theoretical model to explain asymmetry based on internal intertemporal increasing returns, and also used a state space model to obtain some empirical evidence.

The cycle model in this chapter is based on stochastic trigonometric functions, where asymmetry is modelled by specifying the period of the cycle as a function of the steepness. Rather than abruptly switching between two regimes with two distinct cycle periods, the period changes gradually through a continuous range of values. Since our asymmetric cycle specification is a non-linear state space model, basic linear Kalman filter methods are inadequate for estimation. We base our inference on Monte Carlo likelihood estimation and importance sampling techniques.

The remainder of this chapter is organised as follows. In section 5.2 we define the asymmetric UC cycle model, and discuss some of its properties. Section 5.3 contains the State space form and elaborates on estimation methods. An empirical investigation on asymmetries in U.S. macro-economic time series is presented in Section 5.4. Section 5.5 concludes.

5.2 Formulation of asymmetric stochastic cycle components

The basic modelling framework employed here is based on the UC time series model similar as in the previous chapter. Following Beveridge and Nelson (1981), Clark (1989) and Harvey (1989), we assume that many macroeconomic time series can be decomposed into a non-stationary trend $\mu_t$, a stationary cycle $\psi_t$ and an irregular component $\epsilon_t$. The
observed $y_t$ is then modelled as

$$y_t = \mu_t + \psi_t + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, \sigma^2_\epsilon), \quad t = 1, \ldots, n. \quad (5.1)$$

In this section, we focus on the specification of the cyclical component $\psi_t$.

### 5.2.1 Asymmetric deterministic cycles

A deterministic cycle with amplitude $a$, phase $b$ and frequency $\lambda$ can be expressed by

$$\psi_t = a \cos(\lambda t - b), \quad a, b, \lambda, t \in \mathbb{R}, \quad a \neq 0, \lambda \neq 0, \quad (5.2)$$

which is an alternative formulation of the cycle in (4.4).

The frequency $\lambda$ is measured in radians and the period of the cycle is given by $2\pi/\lambda$. The cycle $\psi_t$ is symmetric around its local extrema such that:

$$\psi_{\tau+s} = \psi_{\tau-s}, \quad \psi_{\tau+s} = \psi_{\tau-s} \quad (5.3)$$

for all $s \in \mathbb{R}$ and all $\tau$ for which $\psi_\tau$ is a local minimum or maximum, that is $\partial \psi_t / \partial (\lambda t) |_{t=\tau} = 0$. Since $\dot{\psi}_t = \partial \psi_t / \partial (\lambda t) = -a \sin(\lambda t - b)$, (5.4) it follows that $\lambda \tau = b \pm k\pi$ for $k = 0, 1, 2, \ldots$. We note that the sign of $\dot{\psi}_t$ indicates whether the cycle $\psi_t$ is ascending or descending, while its magnitude determines its steepness.

An asymmetric cycle can be obtained by varying the frequency $\lambda$ for different values of $t$. In the simplest case, the cycle can have different frequencies when $\psi_t$ is ascending or descending. More formally,

$$\psi_t = a \cos(\lambda_t t - b), \quad (5.5)$$

$$\lambda_t = \begin{cases} 
\lambda^+, & \dot{\psi}_t > 0 \\
\lambda^-, & \dot{\psi}_t \leq 0 
\end{cases} \quad (5.6)$$

When $\lambda^+ \neq \lambda^-$, condition (5.3)) does not hold and and resulting cycle is asymmetric but still periodic.

Instead of using two distinct frequencies, we can allow the frequency to depend on a continuous function of $\dot{\psi}_t$, for example,

$$\lambda_t = \lambda + \gamma \dot{\psi}_t, \quad (5.7)$$

specifying the frequency as an affine transformation of the steepness. More generally $\lambda_t$ can be specified as a globally increasing or decreasing function $f(\dot{\psi}_t)$ of the cycle
steepness. However, it is unlikely that very specific forms can be inferred from sparse macro-economic data. We will therefore only consider the simple specification (5.5), (5.7), which captures the asymmetry property in one parameter $\gamma$. For positive values of $\gamma$, the frequency of the cycle is highest when the cycle ascends at its fastest rate, and lowest when it descends at its fastest rate. Figure 5.1 illustrates the two asymmetric cycle specifications, together with their derivatives. Notice that in the first specification the discontinuities in the two regimes are not clearly visible in the cycle, but obvious in the derivative.

We note that our specification explicitly models asymmetry in the steepness of the cycle. Sichel (1993) introduced asymmetry in the deepness, or amplitude of the cycle, while McQueen and Thorley (1993) distinguished asymmetry in roundness, such that positive and negative turning points occur with different acuteness. In the stochastic trigonometric specification, deepness and roundness asymmetry can be incorporated by varying the damping factor and the variance of the disturbance. In this study, we limit the asymmetry to steepness, following the earlier works on asymmetric cycles.

In the formulation of (4.4), the cycles $\psi_t$ in (5.2) and $\dot{\psi}_t$ in (5.4) can be expressed as sine-cosine waves, that is

$$
\psi_t = A \cos(\lambda t) + B \sin(\lambda t), \quad \dot{\psi}_t = B \cos(\lambda t) - A \sin(\lambda t),
$$

Figure 5.1: Stylised asymmetric business cycles with derivatives.

The first plot shows a regime switching cycle, based on one frequency during ascend and one during descend. The second plot shows a smooth frequency evolution, where the frequency is an affine transformation of the cycle slope. The solid line depicts the cycle while the dashed line represents its steepness.
where $A = a \cos b$ and $B = a \sin b$. The reverse transformation is $a = \alpha^2 + \beta^2$ and $b = \tan^{-1}(\beta/\alpha)$. The equivalence follows directly from the first of two trigonometric identities

\begin{align*}
\cos(x \pm y) &= \cos x \cos y \mp \sin x \sin y, \\
\sin(x \pm y) &= \cos x \sin y \pm \sin x \cos y,
\end{align*}

(5.9)

with $x = \lambda t$, $y = b$.

The cycle $\psi_t$ and its partial derivative $\dot{\psi}_t$ can be expressed via a recursion which follows from repeated application of the trigonometric identities (5.9). This recursive expression is given by

\begin{equation}
\begin{pmatrix}
\psi_{t+\delta} \\
\dot{\psi}_{t+\delta}
\end{pmatrix} =
\begin{bmatrix}
\cos(\lambda \delta) & \sin(\lambda \delta) \\
-\sin(\lambda \delta) & \cos(\lambda \delta)
\end{bmatrix}
\begin{pmatrix}
\psi_t \\
\dot{\psi}_t
\end{pmatrix}, \quad \delta > 0, \quad t = 0, \delta, 2\delta, \ldots,
\end{equation}

(5.10)

with $\psi_0 = \alpha$ and $\dot{\psi}_0 = \beta$. The recursion is linear in $\psi_t$ and $\dot{\psi}_t$. A geometric interpretation of (5.10) is obtained by regarding $(\psi_t, \dot{\psi}_t)$ as a point in the Euclidean plane. Multiplication of a point in the plane by the Givens matrix

\begin{equation}
\begin{pmatrix}
\cos(\lambda) & \sin(\lambda) \\
-\sin(\lambda) & \cos(\lambda)
\end{pmatrix}
\end{equation}

(5.11)

rotates the point around the origin. Hence, multiple multiplications moves the point around a circle, while the components $\psi_t$, and $\dot{\psi}_t$ can be seen as projections of the point on the axes of the plane. Cycles based on more general Givens rotations have been considered recently by Luati and Proietti (2009).

The asymmetric cycle (5.7) can be expressed recursively by substituting $\lambda_t$ for $\lambda$ in (5.10). Unlike the expression for the symmetric cycle, this recursion is non-linear in $\psi_t$ and $\dot{\psi}_t$ due to the dependence of $\lambda_t$ on $\dot{\psi}_t$ and the mutual dependence of $\psi_t$ and $\dot{\psi}_t$ for different values of $t$. In the geometrical interpretation, the point $(\psi_t, \dot{\psi}_t)$ rotates around the circle with position dependent velocity.

### 5.2.2 Asymmetric stochastic cycles

A stochastic cycle can be based on (5.10) by including a damping term $\phi$ and random shocks, see Harvey (1989). Similarly we can obtain an asymmetric stochastic cycle but with $\lambda$ in (5.10) replaced by $\lambda_t$ of (5.7) to obtain

\begin{equation}
\begin{pmatrix}
\psi_{t+\delta} \\
\dot{\psi}_{t+\delta}
\end{pmatrix} = \phi
\begin{bmatrix}
\cos(\delta \lambda_t) & \sin(\delta \lambda_t) \\
-\sin(\delta \lambda_t) & \cos(\delta \lambda_t)
\end{bmatrix}
\begin{pmatrix}
\psi_t \\
\dot{\psi}_t
\end{pmatrix} + \begin{pmatrix}
\kappa_t \\
\dot{\kappa}_t
\end{pmatrix},
\end{equation}

(5.12)

\begin{equation}
\lambda_t = \lambda + \gamma \dot{\psi}_t, \quad t = 0, \delta, 2\delta, \ldots,
\end{equation}

(5.13)
where $|\phi| < 1$ is a damping factor, $\lambda_t$ is the time-varying cycle frequency and the disturbance vectors are Gaussian noise:

$$\begin{pmatrix} \kappa_t \\
\dot{\kappa}_t \end{pmatrix} \sim \text{NID} \left(0, \sigma^2_{\kappa} I_2\right), \quad t = 0, \delta, 2\delta, \ldots$$  \hspace{1cm} (5.14)

The damping term $\phi$ ensures that the stochastic process $\psi_t$ is stationary. We note that the frequency $\lambda_t$ is stochastic as a result since $\dot{\psi}_t$ is a stochastic process. In the absence of shocks and with $\phi = 1$, $\psi_t$ and $\dot{\psi}_t$ reduces to the deterministic asymmetric cycle, while a symmetric stochastic cycle is obtained when $\gamma = 0$. In the latter case $\lambda_t = \lambda$, the process $\psi_t$ follows the autoregressive moving average process ARMA(2,1) with the roots of the autoregressive polynomial in the complex range. This property also holds for the process (5.12) conditional on $\lambda_t$. The unconditional process $\psi_t$ follows a non-linear ARMA(2,1) process with the autoregressive coefficients also depending on an ARMA processes.

In the usual discrete time specification of the stochastic cycle the time derivative of $\psi_t$ is not strictly defined, so that it is not immediately obvious that the asymmetry obtained in the deterministic setup carries over. We could construct a continuous time process such that (5.12) is obtained by discretisation. A simpler option is to consider the geometric interpretation. The Givens matrix in (5.12) rotates the point $(\psi_t, \dot{\psi}_t)$ around a circle in a plane, while the shocks $(\kappa_t, \dot{\kappa}_t)$ perturbs the movement after each rotation and the damping factor $\phi$ pulls the point towards the origin to enforce stationarity. Hence in the deterministic rotation dynamics which drives the cyclicality underlying the stochastic cycle $(\psi_t, \dot{\psi}_t)$, $\dot{\psi}_t$ is the slope of the component $\psi_t$ before the random shocks are applied.

5.3 Trend-cycle decomposition: estimation and measurement

5.3.1 Trend-cycle decomposition model

For an observed macro economic time series $y_t$, with $t = 1, \ldots, n$, we consider the model based decomposition given in equation (5.1). In contrast to ARIMA type models, the series are modelled without differencing. Therefore the trend component $\mu_t$ usually requires a non-stationary process.
In our empirical investigation, we employ a smooth trend specification defined by

\begin{align}
\mu_{t+1} &= \mu_t + \beta_t \\
\beta_{t+1} &= \beta_t + \zeta_t \\
\zeta_t &\sim \text{NID}(0, \sigma^2_\zeta), \quad t = 1, \ldots, n, 
\end{align}

where the initial values $\beta_1$ and $\mu_1$ are assumed unknown.

The slope of the trend $\beta_t$ follows a random walk, driven by the disturbance $\zeta_t$. The resulting model for the trend and is also called an integrated random walk.

The cyclical component $\psi_t$ is modelled as an asymmetric stochastic trigonometric process given by (5.12). The cycle is driven by the disturbances $\kappa_t$, $\dot{\kappa}_t$. Similar formulations of asymmetric cycles may be considered for the generalised cycle components of Harvey and Trimbur (2003).

The irregular term $\epsilon_t$ is modelled by Gaussian white noise, $\epsilon_t \sim \text{NID}(0, \sigma^2_\epsilon)$. In many aggregated macro economic series this term is vanishingly small. We assume that the disturbances of the different components are mutually independent, and independent of the initial values of the trend and cycle processes.

### 5.3.2 State space form

The symmetric trend-cycle decomposition model can be cast in a linear Gaussian state space form, with $\alpha_t = \left( \mu_t \quad \beta_t \quad \psi_t \quad \dot{\psi}_t \right)'$ and $\eta_t = \left( 0 \quad \zeta_t \quad \kappa_t \quad \dot{\kappa}_t \right)'$. The system matrices are given by

\begin{align}
T_t &= \begin{bmatrix}
1 & 1 \\
0 & 1 \\
\phi & \begin{bmatrix}
\cos \lambda & \sin \lambda \\
\sin \lambda & -\cos \lambda
\end{bmatrix}
\end{bmatrix}, \\
Z_t &= \begin{bmatrix} 1 & 0 & 1 & 0 \end{bmatrix}, \\
Q_t &= \begin{bmatrix}
0 & 0 & \sigma^2_\zeta \\
0 & \sigma^2_\zeta & 0 \\
\sigma^2_\kappa & 0 & \sigma^2_\dot{\kappa}
\end{bmatrix}, \\
G_t &= \sigma^2_\epsilon,
\end{align}

where $\mathbf{O}$ represents a conforming zero matrix. Since this is a linear Gaussian state space model, state estimation and likelihood evaluation can be handled by standard Kalman filter methods.

When considering the model with an asymmetric cycle $\lambda_t = \lambda + \gamma \dot{\psi}_t$, the model becomes non-linear, with the state transition equation

\[ \alpha_{t+1} = T(\alpha_t) + \eta_t \]
with
\[
T(\alpha_t) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} O \begin{bmatrix} \cos(\lambda_t) & \sin(\lambda_t) \\ -\sin(\lambda_t) & \cos(\lambda_t) \end{bmatrix} \alpha_t, \tag{5.21}
\]
\[
\lambda_t = \lambda + \begin{bmatrix} 0 & 0 & 0 \\ \gamma \end{bmatrix} \alpha_t. \tag{5.22}
\]

### 5.3.3 Importance sampling

The estimation technique adopted to estimate the asymmetric cycle model is the importance sampler of section 3.2. We employ a first order linear approximation of the transition equation to obtain an approximating linear Gaussian model. For the asymmetric cycle model, the transition equation is partly non-linear. The function \(T(\alpha_t)\) needs to be linearised only with respect to \(\dot{\psi}_t\). This implies that only the third and fourth elements of vector \(T(\alpha_t)\) are affected, see (5.21). The third and fourth elements are given by
\[
T_3(\alpha_t) = \phi \cos(\lambda_t) \psi_t + \phi \sin(\lambda_t) \dot{\psi}_t, \tag{5.23}
\]
\[
T_4(\alpha_t) = -\phi \sin(\lambda_t) \psi_t + \phi \cos(\lambda_t) \dot{\psi}_t, \tag{5.24}
\]
respectively. For some given value \((\psi^*_t \ \dot{\psi}^*_t)\) of \((\psi_t \ \dot{\psi}_t)\), the linearisation around \((\psi^*_t \ \dot{\psi}^*_t)\) is given by
\[
T_i(\alpha_t) \approx T_i(\alpha^*_t) + \partial T_i(\alpha_t) / \partial \psi_t \bigg|_{\alpha_t=\alpha^*_t} (\psi_t - \psi^*_t) + \partial T_i(\alpha_t) / \partial \dot{\psi}_t \bigg|_{\alpha_t=\alpha^*_t} (\dot{\psi}_t - \dot{\psi}^*_t), \quad i = 3, 4, \tag{5.25}
\]
where \(\alpha^*_t = (\mu_t \ \beta_t \ \psi^*_t \ \dot{\psi}^*_t)\)', and
\[
\frac{\partial}{\partial \begin{bmatrix} \psi_t \\ \dot{\psi}_t \end{bmatrix}} \begin{bmatrix} T_3(\alpha_t) \\ T_4(\alpha_t) \end{bmatrix} = \begin{bmatrix} \cos(\lambda_t) & \sin(\lambda_t) \\ -\sin(\lambda_t) & \cos(\lambda_t) \end{bmatrix} + \begin{bmatrix} 0 & T_4(\alpha_t) \cdot \partial \lambda_t / \partial \dot{\psi}_t \\ 0 & -T_3(\alpha_t) \cdot \partial \lambda_t / \partial \dot{\psi}_t \end{bmatrix}. \]

For the simple affine transformation we have \(\partial \lambda_t / \partial \dot{\psi}_t = \gamma\). It follows that
\[
\begin{bmatrix} T_3(\alpha_t) \\ T_4(\alpha_t) \end{bmatrix} \approx \begin{bmatrix} T_3(\alpha^*_t) \\ T_4(\alpha^*_t) \end{bmatrix} - R(\alpha^*_t) \begin{bmatrix} \psi^*_t \\ \dot{\psi}^*_t \end{bmatrix} + R(\alpha^*_t) \begin{bmatrix} \psi_t \\ \dot{\psi}_t \end{bmatrix} \tag{5.26}
\]
\[
\approx \gamma \begin{bmatrix} -T_4(\alpha^*_t) \\ T_3(\alpha^*_t) \end{bmatrix} \dot{\psi}^*_t + R(\alpha^*_t) \begin{bmatrix} \psi_t \\ \dot{\psi}_t \end{bmatrix}. \tag{5.27}
\]
This linearised approximation of $T(\alpha_t)$ is used in the non-linear state space model (5.20) to obtain the linearised state space model

$$
\alpha_{t+1} = h_t^* + T_t^* \alpha_t + \eta_t, \quad y_t = Z_t \alpha_t + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, \sigma^2_\epsilon),
$$

(5.28)

where

$$
h_t^* = \begin{bmatrix}
0 \\
0 \\
-\gamma T_4(\alpha_t^*) \dot{\psi}_t^* \\
\gamma T_3(\alpha_t^*) \dot{\psi}_t^*
\end{bmatrix}, \quad T_t^* = \begin{bmatrix}
1 & 1 \\
0 & 1 \\
0 & R(\alpha_t^*)
\end{bmatrix}, \quad t = 1, \ldots, n,
$$

(5.29)

and $Z_t = \begin{bmatrix} 1 & 0 & 1 & 0 \end{bmatrix}$ as before. Note that unlike the symmetric model, the linear approximation of the asymmetric model has time-varying system matrices.

Starting with a trial value for $\alpha_t^*$, repeated evaluation of the first order approximation will converge to the conditional mode of the state of the non-linear model. The converged model is used as the linear approximating model for the importance sampler. The procedure is described, though not implemented in Durbin and Koopman (2001, chapter 11), where examples are given for models with a non-linear observation equation or with non-Gaussian disturbances.

### 5.3.4 Importance weights

In the asymmetric cycle model, the non-linearity only occurs in the transition equation, which implies that the observation density $g(\theta, y|\alpha)$ is equal to $p(\theta, y|\alpha)$. Therefore the importance weights (3.19) simplify to

$$
w_i = \frac{p(\alpha^{(i)}, y)}{g(\alpha^{(i)}, y)} = \frac{p(\alpha^{(i)})}{g(\alpha^{(i)})}
$$

(5.30)

The parts of the densities associated with $\mu_t$ and $\beta_t$ cancel out, since they are identical in the true and approximating model and independent of the cycle. The log-density of the cycle process in the true model, derived from (5.12), (5.14), is given by

$$
\log p(\psi, \dot{\psi}) = C - \frac{1}{2 \sigma_\kappa^2} \sum_t \left( \left( \psi_{t+1} - T_3(\psi_t, \dot{\psi}_t) \right)^2 + \left( \dot{\psi}_{t+1} - T_4(\psi_t, \dot{\psi}_t) \right)^2 \right)
$$

(5.31)

while the log-density of the approximating model for the cycle is given by:

$$
\log g(\psi, \dot{\psi}) = C - \frac{1}{2 \sigma_\kappa^2} \sum_t \left( \left( \psi_{t+1} - \gamma (\dot{\psi}_t - \dot{\psi}_t^*) T_4(\alpha_t^*) - \psi_t \phi \cos \lambda_t^* - \dot{\psi}_t \phi \sin \lambda_t^* \right)^2 \\
+ \left( \dot{\psi}_{t+1} + \gamma (\dot{\psi}_t - \dot{\psi}_t^*) T_3(\alpha_t^*) + \psi_t \phi \sin \lambda_t^* - \dot{\psi}_t \phi \cos \lambda_t^* \right)^2 \right)
$$

(5.32)
The constant term $C$ will cancel in the evaluation of the importance weights. The simulated likelihood can be optimised using numerical maximisation routines. When using a quasi-Newton method, some care must be taken to ensure that the simulated likelihood has a smooth surface. In particular, the same set of random draws for the disturbances must be used when evaluating the likelihood for different values of $\theta$.

## 5.4 Empirical evidence from U.S. economic time series

### 5.4.1 Data description

The empirical relevance of asymmetric cycles is considered for three key time series from the US economy: unemployment (U), gross domestic product (GDP) and gross private domestic investment (IN). The series are obtained from the database of the Federal Reserve Bank of St. Louis at [http://research.stlouisfed.org/fred2](http://research.stlouisfed.org/fred2). The unemployment series is a monthly rate in percentage of civilian unemployment compiled by the Bureau of Labor Statistics. The GDP and investment series are the quarterly chain linked series provided by the Bureau of Economic Analysis. The database-codes of unemployment, GDP and investment are UNRATE, GDP and GDPI, respectively. All three series are seasonally adjusted at the source.

We analysed the three series between 1960 and 2004, using 528 observation for the monthly series and 176 observations for the two quarterly series. The data are plotted in the first panels of figures 5.3, 5.4 and 5.5, along with a smoothed trend estimate. The GDP and investment series can be characterised by a strong trend for the long term while cyclical fluctuations from the trend can be observed. The time series of unemployment is most affected by its cyclical behaviour. It should be noted that both quarterly time series are more cyclical in the 1970s and 1980s than in other years. From the end of the 1980s, the amplitude of the cyclical fluctuations are smaller than in the earlier years. These characteristics in macroeconomic time series have been discussed in Stock and Watson (1993). The monthly time series of unemployment does not have a strong trend and is more subject to typical cyclical dynamics.

### 5.4.2 Parameter estimation for symmetric decomposition model

The basic trend-cycle decomposition model (5.1) with a symmetric cycle is considered first for the quarterly time series GDP and IN and for the monthly time series U. This is a linear Gaussian state space model, and therefore the standard Kalman filter can
be used to compute the likelihood function for a given value of the parameter vector \( \theta \). The likelihood function is maximised with respect to \( \theta \) using numerical optimisation methods. Based on the resulting maximum likelihood estimates, the unobservable trend \( \mu_t \) and cycle \( \psi_t \) can be estimated for \( t = 1, \ldots, n \) (signal extraction) using the Kalman filter and smoother.

The parameter vector \( \theta \) contains the unknown log-variances associated with the irregular, trend and cycle components, \( \log \sigma^2_\epsilon \), \( \log \sigma^2_\zeta \) and \( \log \sigma^2_\kappa \), respectively. The log-variances are estimated so that variances are always positive. The coefficients \( 0 < \rho < 1 \) and \( \omega = 2\pi/\lambda > 2 \) in the cycle model are also included in the parameter vector but transformed by \( \text{lt}(\rho) \) and \( \log(\omega - 2) \) where \( \text{lt}(\cdot) \) is the logistic function, in order to ensure they remain in a valid range. The estimates of \( \theta \) for the trend plus symmetric cycle model are presented in Table 5.1. In the cases of GDP and Inv, the irregular variances are estimated as zero while the trend innovation variance estimates are small. Such estimates are typical for macroeconomic time series, with GDP and IN as examples. These time series have minor irregular changes and are subject to slowly varying (smooth) trends. For the monthly U series, the irregular does exist while the trend is kept smooth.\(^1\) The cycle properties of the three series are quite similar. The persistence is in all cases estimated to be close to unity. The length of the cycle \( \omega \), does differ somewhat. The cycle length for GDP is approximately 6 years while for Inv and Un the length is longer, closer to 9 and 12 years, respectively.

### 5.4.3 Parameter estimation for asymmetric decomposition model

The trend plus asymmetric cycle decomposition model is considered next. The parameter vector \( \theta \) for the previous model is extended with coefficient \( \gamma \) that needs no transformation. The computation of the likelihood function for a given \( \theta \) is carried out by the methods described in the previous section. A linear Gaussian approximating model is constructed as described in section 5.3.4 and samples for the unobserved state vector (with trend \( \mu_t \) and cycle \( \psi_t \)) are generated by the simulation smoothing algorithm. From these samples, the Monte Carlo likelihood function can be evaluated and maximised with respect to \( \theta \). Table 5.1 presents the estimation results of the trend-cycle model for the asymmetric cycles specification, next to the results of the symmetric trend-cycle model. Discussions of the empirical results are given in the next subsection.

In section 3.2.3 we described some suggested testing procedures to validate a central assumption underlying importance sampling, namely the finite value of the importance ration. In table 5.2 the test results are reported for the importance weights for the

\(^1\)A unrestricted estimate of \( \sigma_\zeta \) results in an overly flexible trend, which obscures the cyclical component.
<table>
<thead>
<tr>
<th></th>
<th>U</th>
<th>U (asym)</th>
<th>IN</th>
<th>IN (asym)</th>
<th>GDP</th>
<th>GDP (asym)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_r^2$</td>
<td>$7.70 \times 10^{-4}$</td>
<td>$1.67 \times 10^{-3}$</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>$\sigma_\zeta^2$</td>
<td>$1.13 \times 10^{-7}$</td>
<td>$1.13 \times 10^{-7}$</td>
<td>$1.23 \times 10^{-5}$</td>
<td>$1.21 \times 10^{-6}$</td>
<td>$8.29 \times 10^{-8}$</td>
<td>$7.91 \times 10^{-8}$</td>
</tr>
<tr>
<td>$\sigma_\kappa^2$</td>
<td>$2.77 \times 10^{-2}$</td>
<td>$2.48 \times 10^{-2}$</td>
<td>$2.53 \times 10^{-4}$</td>
<td>$2.44 \times 10^{-4}$</td>
<td>$5.60 \times 10^{-5}$</td>
<td>$5.45 \times 10^{-5}$</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.988</td>
<td>0.989</td>
<td>0.963</td>
<td>0.968</td>
<td>0.950</td>
<td>0.953</td>
</tr>
<tr>
<td>[0.977; 0.993]</td>
<td>[0.979; 0.994]</td>
<td>[0.904; 0.986]</td>
<td>[0.900; 0.990]</td>
<td>[0.898; 0.976]</td>
<td>[0.901; 0.978]</td>
<td></td>
</tr>
<tr>
<td>$\omega$</td>
<td>124.9</td>
<td>102.9</td>
<td>24.0</td>
<td>24.0</td>
<td>36.2</td>
<td>34.8</td>
</tr>
<tr>
<td>[96.6; 161.2]</td>
<td>[82.4; 127.8]</td>
<td>[19.2; 29.9]</td>
<td>[19.3; 29.9]</td>
<td>[26.1; 49.9]</td>
<td>[25.8; 46.7]</td>
<td></td>
</tr>
<tr>
<td>$\gamma$</td>
<td>–</td>
<td>0.00738</td>
<td>–</td>
<td>$-0.36$</td>
<td>–</td>
<td>$-0.91$</td>
</tr>
<tr>
<td>[0.00448; 0.0103]</td>
<td></td>
<td></td>
<td>[-0.64; -0.079]</td>
<td></td>
<td>[-1.70; -0.12]</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.1: Trend cycle decomposition model estimation results.

Maximum likelihood estimates are reported for the trend plus symmetric cycle and trend plus asymmetric cycle (as) model for U.S. unemployment, investment and GDP. Square brackets contain 95% confidence intervals. Normality (N) and Box-Ljung (Q(20)) serial correlation test are also reported, together with log-likelihood values. The likelihood based Wald, LM, LR are asymptotically $\chi^2_1$ distributed.
CHAPTER 5. ASYMMETRIC BUSINESS CYCLES

Table 5.2: Testing the existence of the second moment of importance weights.

Based on the largest of 100,000 generated importance weights, a Pareto distribution is fitted by maximum likelihood. Under the null of finite variance, the asymptotic distributions of the Wald and LM statistics are standard Normal, LR is $0.5(\chi^2_0 + \chi^2_1)$, and the Monahan (M) statistic is negative, and the Pareto parameter $\hat{\nu}$ is less than 0.5. The sample variance of the weights is reported as WgtVar.

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>Wald</th>
<th>LM</th>
<th>LR</th>
<th>WgtVar</th>
<th>$\hat{\nu}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>0.174</td>
<td>2.177</td>
<td>1.299</td>
<td>6.432</td>
<td>159.7</td>
<td>0.619</td>
</tr>
<tr>
<td>IN</td>
<td>-0.060</td>
<td>-1.691</td>
<td>-0.867</td>
<td>3.803</td>
<td>8.487</td>
<td>0.407</td>
</tr>
<tr>
<td>GDP</td>
<td>-0.048</td>
<td>-0.715</td>
<td>-0.418</td>
<td>0.752</td>
<td>14.65</td>
<td>0.461</td>
</tr>
</tbody>
</table>

The tests are calculated for weights generated from 100,000 replications of the state. The tests do not indicate a problem for all series. However, for the unemployment series the existence of a second moment is questionable. This is also evident from the plots of the weights shown in figure 5.2. Model miss-specification is usually the main source of unsatisfactory diagnostics for the importance weights, see the discussion below.

5.4.4 Empirical evidence of asymmetric cycles

First, we note that all three series exhibit asymmetry in the cycle, as is evident from the significant estimate of the $\gamma$ parameter. The symmetric model is a special case of the asymmetric model, with restriction $\gamma = 0$. The LR, Wald and LM test of the validity of the restriction is given in table 5.1. All the statistics indicate that there is significant asymmetry at least at the 5% level. The Unemployment series shows a very large increases in the log-likelihood values. The smallest increase in the log-likelihood is 2.4 points, in the GDP series.

Comparing the symmetric and asymmetric specifications, we observe that in general the cycle disturbances decrease a little, while there is some increase in persistence. For the GDP and Investment series the parameter $\lambda$ changes little between the two specifications. In the Unemployment series, the extracted cycle is quite different in the asymmetric estimates. It is evident that the increased flexibility of the model leads to a different decomposition.

The estimated asymmetry parameter $\gamma$ is positive for Unemployment, which implies short upswings and long downturns. For GDP and Investment the parameter is
5.4. EMPIRICAL EVIDENCE FROM U.S. ECONOMIC TIME SERIES

Figure 5.2: 1,000 largest importance weights in 100,000 simulation draws. Importance weights are ratios of true and approximating densities and used as correction factors to the likelihood of an approximating model. A finite variance of the weights justifies the use of the importance sampling likelihood estimator as the central limit theorem applies.

negative, indicating that periods of growth last longer than those of decline. For Unemployment this result is in line with expectations. In particular, our findings agree with beliefs of classical economists such as Mitchell (1927) and Keynes (1936). However, for output and investment there is less consensus in the literature than for unemployment. For example, Falk (1986)’s application of Nefti’s non-parametric tests of unemployment series did not produce significant results for the U.S. GNP. Clements and Krolzig (2003)’s parametric tests found evidence of asymmetry in the GNP and investment series with a three-regime model, while in a two regime specification the asymmetry was insignificant.

Table 5.1 also includes residual diagnostic tests for serial correlation up to twenty lags (Q(20)) and Normality (N). The Q tests for Investment and GDP are generally satisfactory. The asymmetric specification appears to reduce serial correlation, either the symmetric nor the asymmetric specification show no significance at the 10% level. Normality is rejected for both series, although the asymmetric specification for Investment is a considerable improvement on the symmetric specification. The Normality statistics for the Unemployment series are very large, and Normality is clearly not a good assumption. The asymmetric trend-cycle model does give slightly better results for the residual serial correlation, but this statistic remains significant. The model for
unemployment appears to be inadequate and will likely benefit from a more complete specification for the dynamics.

The last panels of the figures 5.3, 5.4 and 5.5 show that the periods are cyclical, and vary between plausible ranges, generally between five and eleven years. It can also be seen that especially for Investment, the variation in the cyclical component is quite small, especially towards the end. This may account for some difficulties in estimating the likelihood using Monte Carlo methods. It is also evident from the plots that the Unemployment series is quite different in character from the other series: there is no clear direction in it’s trend, and the period of its cycle is large. The magnitude of the cycle is also considerably larger than the those of the other series.

5.5 Conclusion

In this chapter we introduced and estimated an extension to a standard stochastic trigonometric cycle component in UC models to account for asymmetries in the period. Replacing the fixed cycle frequency parameter by an affine transformation of the derivative of the cycle results in a model that can capture the degree of asymmetry by one additional parameter. In contrast to common regime switching specifications, the period varies through a continuous range of values.

The trend plus asymmetric model is presented and estimated in a non-linear state space form. We use a Monte Carlo likelihood approach, where the likelihood is interpreted as an expectation of ratio of densities and estimated by averaging the densities evaluated in simulated values of the unobserved components. In order to obtain a precise estimate with a reasonable number of simulations, importance sampling techniques are used.

The empirical application focuses on three U.S. macro economic time series, unemployment, investment and GDP. We find significant evidence of asymmetry in the three series. The unemployment cycle tends to last longer during declines, while the investment and GDP cycles fall faster than they rise.
5.5. CONCLUSION

Figure 5.3: Trend-cycle decomposition of Unemployment.
The first plot shows the data and smoothed trend, the second plot shows the smoothed asymmetric cycle component, the third plot shows the cycle period.

Figure 5.4: Trend-cycle decomposition of Investment.
The first plot shows the data and smoothed trend, the second plot shows the smoothed asymmetric cycle component, the third plot shows the cycle period.
Figure 5.5: Trend-cycle decomposition of GDP.

The first plot shows the data and smoothed trend, the second plot shows the smoothed asymmetric cycle component, the third plot shows the cycle period.
Chapter 6

Seasonality with Trend and Cycle Interactions

Abstract

Unobserved components time series models decompose a time series into a trend, a season, a cycle, an irregular disturbance, and possibly other components. These models have been successfully applied to many economic time series. The standard assumption of a linear model, often appropriate after a logarithmic transformation of the data, facilitates estimation, testing, forecasting and interpretation. However, in some settings the linear-additive framework may be too restrictive. In this chapter, we formulate a non-linear unobserved components time series model which allows interactions between the trend-cycle component and the seasonal component. The resulting model is cast into a non-linear state space form and estimated by the extended Kalman filter with diffuse initial conditions. We apply our model to U.K. travel data and U.S. unemployment and production series, and show that it can capture increasing seasonal variation and cycle dependent seasonal fluctuations.

6.1 Introduction

A common practice in economic time series analyses and seasonal adjustment procedures is first to take logarithms of the data. Linear Gaussian models can often be fitted to the transformed data, while they are inappropriate for the series in the original metric. The log-additive framework appears to work successfully for time series modelling based on the decomposition in trend, seasonal, irregular and other components. The logarithmic transformation converts an exponentially growing trend into a linear trend. Further, it often eliminates or reduces growing seasonal variation and heteroskedastic-
ity in seasonal time series. However, the log-transformation has various drawbacks. In decomposition models or in seasonal adjustment procedures such as the popular U.S. Census X-11 and X-12 programs, the logarithmic transformation presents a single rigid alternative to the untransformed linear-additive specification, see Findley et al. (1998). In particular, it predicates that time series components combine multiplicatively in the implied model for the untransformed series. A full multiplicative model is not always intended or desired. Moreover, when some heteroskedasticity or changing seasonal variation remains after the transformation, applying the log-transformation again is usually not an attractive solution. Finally, if the data is already supplied in units measuring proportional changes, applying the log-transformation can complicate model interpretation.

In empirical work we may encounter cases where the log-transformation does not remove all heteroskedasticity or growth in the seasonal component. For example, consider the data set of monthly visits abroad by UK residents from January 1980 to December 2006. Figure 6.1 presents time series plots of the data in levels and in logs. The time series of visits abroad shows a clear upwards trend, a pronounced seasonal pattern, and a steady increase of the seasonal variation over time. However, after applying the log-transformation, the increase of seasonal variation has been converted into a decrease. This may indicate that the log transformation is not particularly appropriate for this series. A possible course of action is to consider alternative data transformations.

In this chapter however, we explore a different option by modelling the time series using the class of unobserved components (UC) models, see Harvey (1989) for a detailed
6.2. THE UNOBSERVED COMPONENTS TIME SERIES MODEL

Searching for an appropriate data transformation is essentially a quest for a suitable functional form of the model. Our approach is to alter the functional form directly by relating the seasonal component to other components such as the trend. We introduce a simple multiplicative-additive extension to linear UC models, in which a transformation of the trend acts as a scaling factor to the seasonal component. Estimation is effectively performed using the diffuse extended Kalman filter. Unlike previous studies with multiplicative seasonality in UC models, we explicitly parametrise and estimate the degree of trend-season interaction. The basic linear form is a simple parameter restriction in our model.

When the data contains a cyclical component, the magnitude of the seasonal influence may vary along the phase of the cycle. Although seasonal fluctuations and business cycles are traditionally assumed to be uncorrelated, for some macro-economic series there is increasing evidence that this assumption is not valid. For example Cecchetti et al. (1997), Franses and de Bruin (1999), van Dijk et al. (2003) and Osborn and Matas-Mir (2004) have found varying amounts of interactions between cycles and seasonal adjustment in unemployment and industrial production series using linear or non-linear smooth transition autoregression models. With a straightforward extension of our trend-season interaction model, we also examine interactions between the seasonal component and the business cycle. Interactions between the season and the trend or the cycle are typically studied separately in the literature. The non-linear UC model allows us to model changes in seasonal variation along both trend and cycle fluctuations, as well as changes resulting from exogenous shocks using a single coherent framework.

In the next section, we describe the basic unobserved components model. We further review models with multiplicative seasonality that have been proposed in the literature. In Section 6.3 we introduce our non-linear specification. Empirical applications of the new model are provided in Section 6.4. We conclude with Section 6.5.

6.2 The unobserved components time series model

Unobserved components time series model have proven to be a valuable tools for seasonal adjustment, see for example Kitagawa and Gersch (1984) and Harvey and Scott (1994). Compared to model-free procedures, they offer the benefit of providing statistical tests and prediction algorithms. Additionally, it is simple to incorporate changing seasonal patterns and to introduce additional features such as explanatory variables, interventions and cyclical components. Estimation of parameters and measurement of the components is based on standard Kalman filtering and smoothing methods. In this section we briefly introduce the basic form of the model and provide some details which are needed for the following sections.
The seasonal adjustment framework employed in this chapter is based on the basic structural model (BSM) as described in Harvey (1989). We assume that the time series \( \{ Y_t \} \) is observed which we routinely transform into logs, that is

\[
y_t = \log Y_t, \quad t = 1, \ldots, n.
\]  

(6.1)

The BSM decomposes \( y_t \) into additive stochastic components and is given by

\[
y_t = \mu_t + \gamma_t + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, \sigma^2_{\epsilon}),
\]

(6.2)

where \( \mu_t \) represents the trend, \( \gamma_t \) the seasonal component and \( \epsilon_t \) the irregular disturbance term. The linear model (4.1) can be regarded as a generalisation of the classical time series decomposition in which deterministic components for trend and season are replaced by stochastic processes. The BSM is a simple example of an UC model. It can be extended by including deterministic and/or stochastic components. For example, explanatory variables, intervention effects and stochastic cycles can be a part of the UC model.

The trend component \( \mu_t \) in (6.2) is specified in our applications by the local linear trend model as given by

\[
\begin{align*}
\mu_{t+1} &= \mu_t + \beta_t + \eta_t, \quad \eta_t \sim \text{NID}(0, \sigma^2_{\eta}), \\
\beta_{t+1} &= \beta_t + \zeta_t, \quad \zeta_t \sim \text{NID}(0, \sigma^2_{\zeta}),
\end{align*}
\]

(6.3)

where \( \beta_t \) represents the drift or slope of the trend \( \mu_t \) and the disturbances \( \epsilon_t, \eta_t \) and \( \zeta_t \) are mutually uncorrelated at all lags and leads, for \( t = 1, \ldots, n \). Some notable limiting cases of this specification include: if \( \sigma_{\zeta} \to 0 \) while \( \sigma_{\eta} \) is non-zero the trend is a random walk with drift \( \beta_1 \); if \( \sigma_{\eta} \to 0 \) while \( \sigma_{\zeta} \) is non-zero the trend follows a smooth integrated random walk; when both tend to zero, \( \mu_t \) reverts to a deterministic linear trend. In our empirical section we use a smooth trend specification by restricting \( \sigma^2_{\eta} \) to zero. The initial values of \( \mu_1, \beta_1 \) are generally unknown, and will be represented by a diffuse initial distribution.

The seasonal component \( \gamma_t \) can be specified as a sum of time-varying trigonometric cycles. Specifically, in a model for a time series with seasonal length \( s \), we have

\[
\gamma_t = \sum_{j=1}^{[s/2]} \gamma_{j,t}
\]

(6.4)

where \( [ \cdot ] \) denotes the floor function and the recursion of the components is given by

\[
\begin{pmatrix}
\gamma_{j,t+1} \\
\gamma^*_{j,t+1}
\end{pmatrix} =
\begin{pmatrix}
\cos \lambda_j & \sin \lambda_j \\
-\sin \lambda_j & \cos \lambda_j
\end{pmatrix}
\begin{pmatrix}
\gamma_{j,t} \\
\gamma^*_{j,t}
\end{pmatrix} +
\begin{pmatrix}
\omega_{j,t} \\
\omega^*_{j,t}
\end{pmatrix},
\quad
\begin{pmatrix}
\omega_{j,t} \\
\omega^*_{j,t}
\end{pmatrix} \sim \text{NID}(0, \sigma^2_{\omega} I_2),
\]

(6.5)
6.2. THE UNOBSERVED COMPONENTS TIME SERIES MODEL

with \( \lambda_j = 2\pi j/s \) for \( j = 1, \ldots, \lfloor s/2 \rfloor \) and \( t = 1, \ldots, n \). The seasonal disturbances \( \omega_{jt} \) and \( \omega^*_{jt} \) are uncorrelated with the previously specified disturbances at all lags and leads. Note that the components \( \gamma_{jt+1} \) are not included in \( \gamma_t \) directly, but are used as auxiliary variables to write the seasonal term in recursive form. Further details of the seasonal component can be found in Harvey and Scott (1994) and Proietti (2000) who also describe alternative seasonal component models such as stochastic seasonal dummy variables. Although these alternative specifications can be considered in our non-linear UC model, we restrict ourselves to the trigonometric seasonal component (6.5) in our study. The seasonal components represent non-stationary processes and their initial conditions rely on diffuse distributions, similar to the trend components.

Many macro-economic time series contain periodic fluctuations of a lower frequency than the seasonal frequencies. In economic time series, fluctuations associated with medium frequencies are typically interpreted as the business cycle, as have been discussed in the previous two chapters. The dynamic effects related to these medium frequencies appear often moderately pronounced in the observed economic time series and tend to be of a stationary nature. To incorporate the cyclical dynamics in the time series model, the BSM can be extended by a stochastic cyclical component \( \psi_t \). We then have the decomposition model

\[
y_t = \mu_t + \gamma_t + \psi_t + \epsilon_t, \quad t = 1, \ldots, n, \tag{6.6}
\]

with

\[
\begin{pmatrix}
\psi_{t+1} \\
\psi^*_{t+1}
\end{pmatrix} = \phi \begin{pmatrix}
\cos \lambda c & \sin \lambda c \\
-\sin \lambda c & \cos \lambda c
\end{pmatrix} \begin{pmatrix}
\psi_t \\
\psi^*_t
\end{pmatrix} + \begin{pmatrix}
\kappa_t \\
\kappa^*_t
\end{pmatrix},
\]

\[
\begin{pmatrix}
\kappa_t \\
\kappa^*_t
\end{pmatrix} \sim \text{NID}(0, \sigma^2_\kappa I_2),
\tag{6.7}
\]

where like in the previous chapters, the three unknown coefficients \( \lambda c, \phi \) and \( \sigma^2_\kappa \) in the cycle equation (6.7) represent the cyclical frequency, the damping factor and the cycle disturbance variance, respectively. For \( |\phi| < 1, 0 < \lambda < \pi \), the cycle is a stationary ARMA(2,1) process which collapses into an AR(1) process when \( \lambda c \) approaches zero. For stationary cycle processes represented by (6.7) the unconditional distributions provide the properly defined initial conditions for \( \psi_t \) and \( \psi^*_t \). The disturbances \( \kappa_t \) and \( \kappa^*_t \) are specified to be uncorrelated with the disturbances of the other components at all lags and leads, and uncorrelated with the initial distributions.

The BSM, possibly extended with a cycle component, can be formulated as a linear state space model specified by the equations

\[
y_t = Z \alpha_t + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, \sigma^2_\epsilon),
\]

\[
\alpha_{t+1} = T \alpha_t + \eta_t, \quad \eta_t \sim \text{NID}(0, H), \quad t = 1, \ldots, n, \tag{6.8}
\]
in the notation of (2.1), where we omit time subscripts on the transition and observation matrices and variances, as they are time-homogeneous in the models in this chapter. The system variables $Z, T, H$ depend on unknown parameters, which are the variances of the disturbances in the trend, season and cycle process, and the damping and frequency parameters of the cycle. Together with the observation disturbance variance $\sigma^2_\epsilon$, they can be estimated by maximising the Gaussian likelihood function of the model. After replacing the parameters by their estimated values, the unobserved components can be estimated using the Kalman filtering and smoothing equations. The seasonal adjustment procedure based on BSM simply consists of subtracting the estimated seasonal component $\gamma_t$ from the time series $y_t$, that is $y_{t}^{SA} = y_t - \hat{\gamma}_t$ where $y_{t}^{SA}$ is the seasonally adjusted time series and $\hat{\gamma}_t$ is the estimate of $\gamma_t$ obtained from the Kalman smoothing equations. The filtering, smoothing and likelihood equations for linear Gaussian state space models were provided in chapter 2. Accessible treatments of UC models are provided by Harvey (1993) and Commandeur and Koopman (2007).

### 6.3 Seasonal interacting components

#### 6.3.1 A review of non-linear trend-seasonal models

A mixed additive multiplicative seasonal adjustment procedure based on the classical trend-seasonal-irregular decomposition is considered by Durbin and Murphy (1975) for the modelling of a set of unemployment series. The Durbin-Murphy specification is given by

$$y_t = m_t + g_t + g_t^* m_t + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, \sigma^2_\epsilon), \quad t = 1, \ldots, n,$$

where $m_t$ is a deterministic trend function, $g_t$ is an additive seasonal fixed effect and $g_t^*$ is a multiplicative seasonal fixed factor. A standard moving average filter can be derived to extract the different components from the data. Although this model was not based on a stochastic UC model, it can be regarded as an early precursor to our multiplicative seasonal component model presented in Section 6.3.2 below.

Bowerman et al. (1990) explored a number of different approaches to deal with the problem of increasing seasonal variation in time series. Although all their suggestions are built on autoregressive moving average model based methods, one of their models takes a similar direction to what we propose. In their seasonal interaction model, changes in the seasonal component are directly related to a deterministic trend by

$$y_t = \beta_0 + \beta_0^+ t + \sum_{j=1}^{s-1} \beta_j D_{j,t} + \sum_{j=1}^{s-1} \beta_j^+ D_{j,t} t + u_t,$$
where $\beta_0 + \beta^+_0 t$ is the fixed trend component with unknown coefficients $\beta_0$ and $\beta^+_0$, $D_{j,t}$ is the seasonal dummy regression variable with unknown coefficients $\beta_j$ and $\beta^+_j$ for $j = 1, \ldots, s - 1$ and $u_t$ is modelled as an autoregressive integrated moving average process. The coefficients $\beta_j$ are associated with the seasonal effects that are independent of the trend while the coefficients $\beta^+_j$ are interacting with the trend.

In most current applications of UC models, the specifications are of the logarithmic additive type, which can be easily formulated as a linear state space model. An important practical advantage of linearity is that optimal estimates of the latent components, parameters and model predictions are easily obtained using standard Kalman filter based methods. Estimation is a routine procedure for which user-friendly graphical packages are available. Combining multiplicative and additive components may result in a better model fit. However, optimal estimation in such models can be quite complex, and are often carried out using elaborate and computationally expensive simulation methods. For example, Shephard (1994) formulated the multiplicative UC seasonal adjustment model

$$y_t = (1 + \gamma_t + \epsilon^+_t)\mu_t + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, \sigma^2_\epsilon), \quad \epsilon^+_t \sim \text{NID}(0, \sigma^2_{\epsilon^+}),$$

for $t = 1, \ldots, n$, where $\mu_t$ and $\gamma_t$ are the trend and seasonal components and can possibly be modelled as in (6.3) and (6.5), respectively. The two irregular terms $\epsilon_t$ and $\epsilon^+_t$ are uncorrelated with each other and with all other disturbances at all leads and lags. The seasonal term $\gamma_t$ interacts with the trend component through scaling while the irregular term $\epsilon^+_t$ allows for additional heteroskedasticity. The multiplicative UC model was used to seasonally adjust the U.K. M4 money supply series based on parameter estimates obtained from Markov chain Monte Carlo methods. Durbin and Koopman (2001) used a similar additive-multiplicative specification as an exposition example for importance sampling techniques.

An alternative specification, called the pseudo-additive decomposition is described in Findley et al. (1998). It was developed by the U.K. Office for National Statistics for seasonally adjusting series with very small, possibly zero values in certain seasons. The decomposition is based upon the representation

$$Y_t = T_t(S_t + I_t - 1)$$

where $Y_t$ is the time series in levels while $T_t, S_t$ and $I_t$ are the trend, seasonal and irregular components, respectively. The main feature of this decomposition is that the seasonal adjustment is carried out by substracting the term $T_t(S_t - 1)$ rather than division by a seasonal factor which is unstable when the seasonal factor is very small. Moving average based filters to extract the components are implemented in the X-12-ARIMA program of the U.S. Bureau of the Census.
Proietti and Riani (2009) consider the use of the Box-Cox transformation in seasonal UC models as a generalisation of the log-transformation. This approach implies an inverse Box-Cox transformation on the sum of the components and it allows for a far wider range of options than the usual exponential transformation. However, interpretation in the original metric can be awkward for many values of the Box-Cox transformation parameter. The model was estimated with a combination of numerical integration and simulation techniques.

Finally, the methodology of Ozaki and Thomson (1994) is close to our non-linear UC model of Section 6.3.2 below although the specifics and motivations of the models are different. Ozaki and Thompson consider a UC model in levels, given by

$$Y_t = M_t(1 + G_t)e^{\epsilon_t - \sigma^2/2}, \quad \epsilon_t \sim \text{NID}(0, \sigma^2), \quad t = 1, \ldots, n,$$

where $M_t$ is a linear Gaussian stochastic process for the trend while $G_t$ is a stochastic seasonal component. When the log-transformation is applied to $Y_t$, the model for $y_t = \log Y_t$ becomes linear and is given by

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

for $t = 1, \ldots, n$. Parameter estimation is carried out on basis of the non-linear model for $Y_t$ using the extended Kalman filter rather than fitting the linear model to the log-transformed series $y_t$. The main motivation of this approach is to provide a model-based framework for the X-11 seasonal adjustment procedure.

### 6.3.2 Trend and cycle interactions in the basic structural model

The standard linear BSM of Section 6.2 is usually fitted to log-transformed data, implying a model with multiplicative components in the untransformed series. In the previous section we have discussed a number of alternative specifications that have been suggested in the literature. These non-linear model specifications can be considered when heteroskedasticity or changing seasonal variation is not adequately removed by the model-based seasonal adjustment procedure. We propose to generalise the BSM by scaling the amplitude of the seasonal component via an exponential transformation of the trend component. The time series, either in levels or in logs, is decomposed by the non-linear model

$$y_t = \mu_t + e^{b_{\mu_t}} \gamma_t + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, \sigma^2), \quad t = 1, \ldots, n,$$

where $b$ is an unknown fixed coefficient while the dynamic specification of the trend component $\mu_t$ is given by (6.3) and the seasonal component $\gamma_t$ is given by (6.5). The sign of the coefficient $b$ determines whether the seasonal variation increases or decreases when
a positive change in the trend occurs. The model reduces to the basic linear specification when $b$ is zero. The overall amplitude of the seasonal component is determined by both $e^{b\mu_t}$ and the disturbance variance $\sigma^2_\omega$ in the stochastic seasonal equation (6.5). The two sources of seasonal amplitude can be made more explicit by restricting $\sigma^2_\omega = 1$ in (6.5) and replacing $e^{b\mu_t}$ by $e^{a+b\mu_t}$ as the scaling process in (6.9) where $a$ is a fixed unknown coefficient. However, we adopt the specification in (6.9) to remain close to the original linear BSM.

When the cycle component (6.7) is added to the BSM we obtain model (6.6). Similar to the introduction of the trend interaction, we can extend model (6.6) by a trend and cycle interaction to obtain the non-linear model

$$y_t = \mu_t + \psi_t + e^{b\mu_t + c\psi_t}\gamma_t + \epsilon_t,$$

(6.10)

where $c$ is an unknown fixed coefficient. The seasonal term in (6.10) is scaled by an exponential transformation of a linear combination of the trend and cycle components. In economic time series, the $\psi_t$ component can often be referred to as the business cycle. In this case, the sign of $c$ determines whether seasonal effects are amplified or dampened during expansions and recessions. The restriction $b = c$ implies that the seasonal component is scaled by the combined trend-cycle component $\mu_t + \psi_t$. Model (6.10) reduces to model (6.6) when $b = c = 0$.

In specification (6.10) changes to the seasonal pattern can be due to either the random shocks of $\omega_{j,t}$ and $\omega_{j,t}^*$ in (6.5) or to changes in the trend and cycle components. It is possible to generalise the trend-cycle interaction model further. For instance, we can introduce a scaling process to the cyclical component based on the trend and seasonal components. We can also include interactions based on exogenous intervention and regression variables. In this study however we limit ourselves to the specifications described in this section.

### 6.3.3 Seasonal interaction model in state space form

The non-linear seasonal interaction model cannot be formulated in the linear state space form (6.8). Therefore we consider a non-linear state space model where the observation equation $y_t = Z\alpha_t + \epsilon_t$ in (6.8) is replaced by

$$y_t = Z(\alpha_t) + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, \sigma^2_\epsilon), \quad t = 1, \ldots, n.$$

(6.11)

The trend and cycle interaction model (6.10) has a state space representation with a state vector given by

$$\alpha_t = (\mu_t \quad \beta_t \quad \psi_t \quad \psi^*_t \quad \gamma_{1,t} \quad \gamma^*_1 \quad \gamma_{2,t} \quad \gamma^*_2 \quad \cdots \quad \gamma_{\lfloor s/2 \rfloor,t} \quad \gamma^*_{\lfloor s/2 \rfloor,t})',$n

(6.12)
for $s$ is even and with the non-linear equation $Z(\alpha_t)$ in (6.11) given by

$$Z(\alpha_t) = \mu_t + \psi_t + e^{b\mu_t+c\psi_t}\gamma_t,$$

(6.13)

for $t = 1, \ldots, n$. The dynamic specifications of the components are formulated in the state equation of (6.8) with system variables given by

$$T = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & T^\psi & 0 \\ 0 & 0 & 0 & T^\gamma \end{bmatrix}, \quad H = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \sigma^2_\xi & 0 & 0 \\ 0 & 0 & \sigma^2_\kappa I_2 & 0 \\ 0 & 0 & 0 & \sigma^2_\omega I_{s-1} \end{bmatrix},$$

(6.14)

where

$$T^\psi = \phi C(\lambda^c), \quad T^\gamma = \text{diag}\{ C(\lambda_1) \ldots C(\lambda_{\lfloor s/2 \rfloor}) \ldots \text{1} \},$$

(6.15)

for $s$ is even and with $\lambda_j = \frac{2\pi j}{s}$ for $j = 1, \ldots, \lfloor s/2 \rfloor$ and

$$C(\lambda) = \begin{bmatrix} \cos \lambda & \sin \lambda \\ -\sin \lambda & \cos \lambda \end{bmatrix}.$$
6.4 APPLICATIONS

descriptions of the real world dynamics that generated the data. The non-linear model formulations can be adopted to improve the model fit compared to existing linear specifications. The inaccuracy due to the linearisation step in the extended filter does not need to be interpreted as an error. When the extended filter yields a reasonable fit to the data, we could interpret the linearisation as the desired model, and the non-linear model as an auxiliary device to formulate it. Although non-linear models are easier to formulate and to explicate than their linearisations, there are usually few theoretical arguments to suppose that they are necessarily a better approximation to the underlying data generating process. Nor does a linearisation need to be superior from a forecasting point of view. Finally, Gaussianity assumptions on disturbance terms in economic models have more often than not practical rather than theoretical motivations. Hence, re-assuming conditional Gaussian state distributions in each iteration step as the extended filter effectively does (if we follow a derivation based on Gaussian disturbances), is not necessarily inferior as a description of reality than the assumption implied by assuming Gaussian disturbances in the non-linear form.

6.4 Applications

6.4.1 U.K. visits abroad

We first consider the data set of monthly visits abroad by U.K. residents from January 1980 to December 2006 that is introduced in Section 1. The data is compiled by the Office for National Statistics (ONS), based on the International Passenger Survey. Time series plots of the series in levels and in logs were presented in Figure 6.1 in the introduction of this chapter. A linear UC model with smooth trend, trigonometric seasonal and cycle components together with a Normal white noise disturbance term as given by equations (6.3), (6.5), (6.6) and (6.7) is considered first for the number of visitors in levels (millions of visits). The maximum likelihood estimates of the parameters are

\[
\hat{\sigma}_\epsilon = 0.106, \quad \hat{\sigma}_\zeta = 0.00062, \quad \hat{\sigma}_\omega = 0.0119, \\
\hat{\sigma}_\kappa = 0.00050, \quad \hat{\phi} = 0.958, \quad 2\pi/\lambda^c = 123, \quad \log L = 48.2, \tag{6.18}
\]

where \( \log L \) is the log-likelihood value of the model evaluated at the maximum likelihood estimates of the parameters. The estimated standard deviation of the trend disturbance is relatively small, which implies that the trend is quite steady. The cycle has the period \( 2\pi/\lambda^c \) which is estimated by 123 months. Furthermore, the estimates of the cycle parameters include a relatively small disturbance and amplitude. Most of the variation in the series can be attributed to the seasonal component and to the disturbance term.
Alongside estimates of the unobserved state variables, the (extended) Kalman filter provides one-step ahead prediction errors $v_t$ and their variances $f_t$ in each step of the recursion. Diagnostic tests can be based on the standardised prediction errors $v_t / \sqrt{f_t}$. Here, and in the following examples, we calculate the test statistics

$$N(\chi^2_2) = 6.21, \quad H_{104}(F_{104,104}) = 2.46, \quad Q_{12}(\chi^2_{11}) = 30.9, \quad Q_{24}(\chi^2_{23}) = 43.8, \quad (6.19)$$

where $N$, $H$ and $Q_l$ denote statistics for Normality, heteroskedasticity and serial correlation up to lag $l$, respectively. The null distributions of the tests are given between parentheses. The Normality statistic that we use is based on the sample skewness and kurtosis, and described in detail in Doornik and Hansen (1994). The statistic is a modification of the well known test by Bowman and Shenton (1975), with alternative distributional assumptions which are more suitable for small samples. The heteroskedasticity statistic is the classic Goldfeld-Quandt test. As suggested by Goldfeld and Quandt (1965), we do not use the entire sample in order to increase the power of the test. In all our examples, we only test for different variances between the first and last one third of the series. Finally the serial correlation statistic is the standard Ljung-Box portmanteau test (Ljung and Box, 1978).

The diagnostics indicate that there is significant residual heteroskedasticity and serial correlation in the estimated linear model for the passenger survey series, as each is significant at the 5% level.

Next the non-linear specification

$$y_t = \mu_t + \psi_t + e^{b\mu_t} \gamma_t + \epsilon_t, \quad (6.20)$$

is considered, for which parameter estimates are obtained by applying the EKF from Section 6.3.2. The parameter and likelihood estimates are given by

$$\hat{\sigma}_\epsilon = 0.116, \quad \hat{\sigma}_\zeta = 0.00090, \quad \hat{\sigma}_\omega = 0.00611, \quad \hat{b} = 0.0984$$

$$\hat{\sigma}_\kappa = 0.00088, \quad \hat{\phi} = 0.921, \quad 2\pi/\hat{\lambda}_c = 589, \quad \log L = 55.1. \quad (6.21)$$

The most striking difference between the estimates of the linear model and model (6.20) is the large drop in the value of the standard deviation of the seasonal disturbances, from 0.0119 to 0.00611. The seasonal component $\gamma_t$ in the non-linear model is scaled by the process $e^{b\mu_t}$, which must account for most of the drop in $\sigma_\omega$. The variation in $\mu_t$ is now factored in the seasonality through the scaling in $e^{b\mu_t}\gamma_t$. The process for $\gamma_t$ itself fluctuates less as a result. The upper graph of Figure 6.2 illustrates this by showing the scaled ($e^{b\mu_t}\gamma_t$) and unscaled ($\gamma_t$) seasonal components as estimated by the extended Kalman smoother. The scaled component is changing largely due to the trend component, while the unscaled component shows much smaller movements and
6.4. APPLICATIONS

Figure 6.2: Visits of U.K. residents abroad: (i) smooth estimates of the scaled and unscaled seasonal components obtained by the EKF and its associated smoothing equations; (ii) scaling process $e^{b\mu_t}$ with $\mu_t$ replaced by its smoothed estimate.

consequently does not require a large standard deviation in the disturbance. We confirm that the scaled component is roughly the same as the estimated $\gamma_t$ from the linear model. In the non-linear model, the cycle frequency $\lambda^c$ approaches zero, which implies that the period approaches infinity and the cycle process $\psi_t$ reduces to a first order autoregressive process. Relative to the sample size of the series, the estimated period of the cycle component was already quite large for the linear model. It highlights the difficulty in empirically identifying the cycle component from this data set accurately. We do not consider this a major problem, since relative to the other components the impact of the cycle is very modest in this series.

The diagnostic tests for the residuals of the non-linear model are given by

$$N(\chi^2_2) = 3.18, \quad H_{104}(F_{104,104}) = 1.85, \quad Q_{12}(\chi^2_{11}) = 21.9, \quad Q_{24}(\chi^2_{23}) = 31.0. \quad (6.22)$$

Compared to the previous linear model, all the diagnostic tests have improved. The $Q_{12}$ and the $H$ statistic are still significant at the 5% level, but the statistics indicate that autocorrelation and heteroskedasticity are less severe than they were in the initial specification. Taken together with the significant increase in the log-likelihood, we conclude that the non-linear model is a clear improvement over the linear specification.
6.4.2 U.S. unemployment

In this section we apply the seasonal interaction model to the log of the number of unemployed people in the U.S. The monthly data set was obtained from the U.S. Bureau of Labor Statistics and spans the period from January 1948 to December 2006.

A graph of the log-unemployment with the estimated trend from a linear decomposition model with trend, season, cycle and irregular components is shown in Figure 6.3. Salient features in the series are an overall increasing trend that levels off towards the end, a distinct seasonal pattern, and a large amount of medium frequency cyclical fluctuation. The estimated parameters from the linear decomposition model considered are given by

\[
\hat{\sigma}_\epsilon = 0.00034, \quad \hat{\sigma}_\zeta = 0.00150, \quad \hat{\sigma}_\omega = 0.00120, \\
\hat{\sigma}_\kappa = 0.00072, \quad \hat{\phi} = 0.970, \quad 2\pi/\hat{\lambda}^c = 57, \quad \log L = 1082.3, \\
\]

with diagnostics

\[
N(\chi^2_2) = 70.6, \quad H_{228}(F_{228,228}) = 4.43, \quad Q_{12}(\chi^2_{11}) = 18.5, \quad Q_{24}(\chi^2_{23}) = 33.9. \quad (6.23)
\]

The business cycle is quite persistent, with a damping factor of 0.97 for a monthly frequency, which corresponds to 0.7 for a yearly frequency. The period of the cycle is close to five years, which is a typical business cycle frequency. In Figure 6.3 the estimated trend is displayed, and it may be concluded that the series possibly contains a second
cycle with a longer period that is currently captured by the trend component. The prediction error based diagnostic tests indicate that Normality and homoskedasticity are strongly rejected, while the serial correlation statistics are not significant at the 5% level.

The non-linear model with interactions between the trend plus cycle and the seasonal component is given by equation (6.10) and is considered next. First we concentrate on the cycle-season interaction and estimated the parameters of this model under the constraint $b = 0$. Maximum likelihood estimates are given by

$$
\hat{\sigma}_e = 0.00034, \quad \hat{\sigma}_\zeta = 0.00217, \quad \hat{\phi} = -0.58 \quad \hat{\sigma}_\omega = 0.00117, \\
\hat{\sigma}_\kappa = 0.00065, \quad \phi = 0.968, \quad 2\pi/\hat{\lambda}^c = 53, \quad \log L = 1098.0,
$$

(6.25)

with diagnostics

$$
N(\chi^2_2) = 47.3, \quad H_{228}(F_{228,228}) = 4.22, \quad Q_{12}(\chi^2_{11}) = 15.0, \quad Q_{24}(\chi^2_{23}) = 32.5.
$$

(6.26)

Compared to the estimates from the linear model, the cycle length becomes slightly shorter. The model fit has greatly improved in terms of the increase in the log-likelihood value, which gains almost 16 points at the cost of only one parameter. The diagnostic tests show small improvements, but the Normality and heteroskedasticity tests remain highly significant. The negative value of the estimated coefficient $\hat{c}$ indicates that the seasonal component is dampened during periods of high cyclical unemployment and attenuated in the negative phases of the cycle, which is consistent with the findings of Franses (1995). The smoothed scaling process is depicted in Figure 6.4, together with the scaled and unscaled seasonal component. The plot shows that the scaling process adds about 20% cyclical variation to $\gamma_t$ in the early parts of the series, but levels off towards the end as the amplitude of the estimated cycle component wanes in the last decades.

We estimate a non-linear model with both trend-season and cycle-season interactions next by relaxing the restriction $b = 0$. The parameters estimates are given by

$$
\hat{\sigma}_e = 0.00033, \quad \hat{\sigma}_\zeta = 0.00050, \quad \hat{b} = -0.024, \quad \hat{c} = -0.53 \quad \hat{\sigma}_\omega = 0.00126, \\
\hat{\sigma}_\kappa = 0.00074, \quad \phi = 0.980, \quad 2\pi/\hat{\lambda}^c = 76, \quad \log L = 1106.9,
$$

(6.27)

with diagnostics

$$
N(\chi^2_2) = 49.2, \quad H_{228}(F_{228,228}) = 4.47, \quad Q_{12}(\chi^2_{11}) = 23.3, \quad Q_{24}(\chi^2_{23}) = 45.9.
$$

(6.28)

Although there is a considerable increase in the likelihood, not all the diagnostic statistics have improved. The most notable difference is that the serial correlation in the residuals is more severe than it was in the original linear specification, and both $Q$ tests
are now significant at the 5% level. This may be attributed to the fact that compared to previous estimates, there is a considerable change in the decomposition, as the trend is smoother, while the cycle period has lengthened over six years. Thus, a direct comparison with the linear specification is difficult, in contrast to the previous non-linear specification without a trend-season interaction. If we fix the cycle length at the value of the previous model and re-estimate, we obtain the results

$$\hat{\sigma}_\epsilon = 0.00034, \quad \hat{\sigma}_\zeta = 0.00143, \quad \hat{b} = -0.021, \quad \hat{c} = -0.60 \quad \hat{\sigma}_\omega = 0.00132,$$

$$\hat{\sigma}_\kappa = 0.00067, \quad \hat{\phi} = 0.973, \quad 2\pi/\lambda^c = 53, \quad \log L = 1104.6,$$

and diagnostics

$$N(\chi^2_2) = 49.4, \quad H_{228}(F_{228,228}) = 4.04, \quad Q_{12}(\chi^2_{11}) = 17.6, \quad Q_{24}(\chi^2_{23}) = 34.2. \quad (6.29)$$

These diagnostic test statistics are very close to those obtained from the model with only cycle-season interaction. The $Q$ statistics have fallen below the 5% significance level again, while the model retains most of the improvement in the likelihood of the previous model. Thus, we prefer this decomposition to the previous one with an unrestricted cycle frequency parameter. Nevertheless, the diagnostics are still not quite satisfactory for this series, as non-Normality and heteroskedasticity remain severe. As noted by Franses (1995), the U.S. unemployment series exhibits several types of non-linearities that we have not incorporated in our model, such as asymmetry in the cycle and different shock persistence in different regimes, which suggests possible directions to improve upon the current specification.
6.4. APPLICATIONS

6.4.3 U.S. industrial and dwellings production

In our final empirical application we consider the seasonal interaction model for the U.S. industry and dwellings production series, obtained from OECD Main Economic Indicators 2007 release 07. Both monthly series start in January 1960 and end in December 2006. The production of total industry is an index series standardised at 100 in the year 2000, while the production of dwellings is measured in billion U.S. dollars. We model both series in logarithms.

Table 6.1 presents the estimation results of linear and non-linear models for both series. For the industrial production series, we allow both trend-season and cycle-season interactions in the non-linear model. The estimates show that there is almost no improvement resulting from using the more general non-linear specification, as the log-likelihood value of the non-linear model increases by only 0.6. Since the linear model is a simple dimensional reduction in the parameter space of the non-linear model, we can use a Likelihood Ratio (LR) test to assess the significance of the linearity restriction. Under the null hypothesis that the parameter restrictions are valid, the LR statistic is asymptotically $\chi^2_k$ distributed. In this example, the statistic evaluates to 1.2, which would not be significant with two parameter restrictions under any conventional significance level. We therefore conclude that no trend or cycle induced variations in the seasonal component of the U.S. industrial production series is detected by our model.

For the dwelling production series, we estimate the parameters for non-linear models with only a trend-season interaction ($c = 0$), only a cycle-season interaction ($b = 0$) and with both trend-season and cycle-season interactions. We learn from Table 6.1 that the models with only the trend-season interaction and only the cycle-season interaction improve the fit compared to the linear model. The model with both interactions provides a better fit than the two models with single interactions. Using the LR test and starting from the unrestricted non-linear model, we can reject the restrictions $b = 0$ and $c = 0$ individually at the 5% significance level. The joint restriction $b = c = 0$ can also be rejected at the 5% level. The fit of each interaction model provides a clear improvement when compared to the fit of the linear model. The residual diagnostic tests are all significant while those of the non-linear models show only small improvements. This suggests that for this series a more elaborate modelling effort may be fruitful.

The estimated coefficient of the trend-season interaction is negative, which implies that the seasonal variation decreases with an increase in the trend. It can be argued that technological changes which may have reduced seasonal dependence in dwellings productions in the past decades have coincided with the trending behaviour in the series, which is likely caused by many unrelated factors such as demographic trends or changing preferences. Our model does not distinguish between underlying causes, but
Table 6.1: U.S. monthly industrial and dwellings production, linear and non-linear model estimates and test statistics.

<table>
<thead>
<tr>
<th>Year</th>
<th>Linear Trend</th>
<th>Linear Cycle</th>
<th>Non-linear Trend</th>
<th>Non-linear Cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>-0.0007</td>
<td>0.0007</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>1981</td>
<td>-0.0007</td>
<td>0.0007</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>1982</td>
<td>-0.0007</td>
<td>0.0007</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>1983</td>
<td>-0.0007</td>
<td>0.0007</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>1984</td>
<td>-0.0007</td>
<td>0.0007</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>1985</td>
<td>-0.0007</td>
<td>0.0007</td>
<td>0.001</td>
<td>0.001</td>
</tr>
</tbody>
</table>

The log likelihoods for the linear and non-linear models are as follows:

- Linear model: $\log L = 1742.0, 1742.6, 1742.7, 1742.0, 1742.0, 1742.0$
- Non-linear model: $\log L = 1182.7, 1186.0, 1187.5, 1187.5, 1187.5, 1187.5$

Test statistics:

- $H = 2.18, 2.16, 2.49, 2.48, 2.40, 2.41$
- $Q = 34.7, 35.2, 130.7, 128.1, 120.6, 120.5$
- $Q = 76.7, 78.7, 218, 218, 218, 218$

The lag length ($\ell$) is determined by maximizing the log-likelihood function.
merely reflects the effect of permitting the interaction.

The negative coefficient of the cycle-season interaction indicates that the seasonality in the dwellings productions is contra-cyclical, that is, the seasonal amplitude decreases with upswings in the production. A similar effect has been documented in some other U.S. industries by Cecchetti et al. (1997), who interpret it as a capacity constraint in the sector when the inventory series of the sector does not show a decrease. However, we do not use inventory series in our investigation.

Figure 6.5 shows the estimated scaling process \( \exp(b \mu_t + c \psi_t) \) from all three non-linear models. We observe that the trend induced reduction in the seasonality is fairly uniform and is roughly 20% over the sample period. The cyclical swings contribute about 10% at their top in the mid 1970s and early 1980s. The cycle induced variations in seasonality seems to have reduced since the early 1990s, as a direct result of the declining amplitude of the cyclical component. Finally, we present the unobserved components decomposition of the non-linear model with both trend-season and cycle-season interactions in Figure 6.6. The figure includes the data and estimated trend, cycle scaled and unscaled seasonal component.
6.5 Conclusion

In this chapter we have presented a simple non-linear extension to the basic unobserved components model to allow the seasonal term to interact with the trend and cycle components. The model that we propose addresses some functional miss-specifications that may arise from imposing a (transformed) additive components structure. Starting from a basic linear unobserved components model with trend, season, cycle and irregular components, we include a transformed linear combination of trend and cycle components as a scaling factor for the seasonal component. In the resulting model, the seasonal amplitude is amplified or dampened along movements of the trend and cycle, depending on the estimated parameters.

In our empirical applications, we have considered models for U.K. travel, U.S. unemployment and U.S. production data. The travel data contains increasing seasonal variation, which is not adequately removed by a logarithmic transformation. Our non-linear model shows a significant improvement in the model fit and provides better residual diagnostics. In the unemployment series, we found significant interactions between the cycle and the seasonal term. Although the model improves on the linear specification, it does not capture all the non-linear dynamics in the series. The estimated coefficient sign indicates that seasonal effects are dampened during recessions. Finally our parameter estimates for the U.S. production series do not show evidence of interactions between
in the total productions series. However, in the production of dwellings, we observe a significant contra-cyclical effect, as well as a dampening of seasonal fluctuations along the increasing trend.
CHAPTER 6. SEASONALITY WITH TREND AND CYCLE INTERACTIONS
Chapter 7

Estimating Stochastic Volatility Models using Importance Sampling

Abstract:
The variance of disturbances in many time series are not constant. Often clusters of high and low variances can be observed. The stochastic volatility model is a widely used model to describe such behaviour, especially in the finance literature. Parameter estimation in stochastic volatility models is not straightforward. In this chapter, we describe and compare two simulated importance sampling based maximum likelihood estimators for a basic stochastic volatility model. Their effectiveness and estimation performance are assessed using a simulation study. Further, we validate the two methods using diagnostic importance sampling test procedures. Stochastic volatility models with both Gaussian and Student-t distributed disturbances are considered.

7.1 Introduction

Modelling price volatility gained much interest in the finance and econometrics literature over the past decades. In empirical econometrics the dominant models volatility models have been generalizations and extensions of the autoregressive conditional heteroskedasticity (ARCH) specification introduced by Engle (1982), while in the theoretical finance literature, stochastic volatility in continuous time are the most prominent models.

As financial asset price volatility is not directly observable, various measurement methods have been devised. One widely used approach is to assume a model for a contingent asset (usually a traded financial option) with an observable price which depends on the volatility of the original asset. The volatility measure of the original asset
inferred from the contingent asset price is usually referred to as the implied volatility (Latané and Rendleman, 1976; Chiras and Manaster, 1978). In this chapter we consider discrete time stochastic volatility (SV) models, which do not depend on prices of other assets, but attempt to estimate the volatility directly from the original asset price. The volatility is modelled as a latent stochastic process in the state space form. The model is typically applied to day-to-day currency or asset price returns or . A more recent development is the realised variance measure, which is computed from high frequency intra-day data (Andersen et al., 2001; Barndorff-Nielsen and Shephard, 2002; Andersen et al., 2003). These are typically calculated by summing five or ten minute squared returns over each day. Such measures can be expected to be more informative than estimates based on day-to-day returns. However, acquisition and management of the large data collections that are required for high frequency data are still relatively costly.

In the next section we describe the basic model specification and give a brief overview of estimation methods and problems. Section 7.3 and 7.4 provide details of importance sampling for SV models. The performance of two specific samplers are empirically compared using simulated data in section 7.5 and 7.6 The SV model is adapted to incorporate Student-t distributed disturbances in section 6.7. Section 7.8 concludes the chapter.

7.2 Model specification and estimation

The basic discrete time SV model described in Taylor (1986), can be written as

\[ y_t = e^{h_t/2} \epsilon_t, \]  
\[ h_{t+1} = \phi_0 + \phi_1 h_t + \sigma \eta_t, \]  
\[ h_1 \sim N \left( \frac{\phi_0}{1 - \phi_1}, \frac{\sigma^2}{1 - \phi_1^2} \right), \]

for \( t = 1, \ldots, n \). The disturbance processes \( \{\epsilon_t\} \) and \( \{\eta_t\} \) are assumed to be standard Gaussian, serially and mutually uncorrelated at all lags and leads, and uncorrelated with \( h_1 \). The stationary AR(1) process \( \{h_t\} \) is not observed, and represents the log-variance of the observations \( \{y_t\} \). This model has three unknown parameters \( (\phi_0, \phi_1, \sigma^2) \), with restrictions \( 0 \leq \phi_1 < 1 \) and \( \sigma^2 > 0 \). In section 7.6, the model is generalised by considering a t distribution for \( \epsilon_t \).

Conditional on \( \{h_t\} \), the observation process is simply heteroskedastic Gaussian noise. The model a discrete time approximation of the continuous time SV models of Hull and White (1987). For an overview of the statistical properties and the theoretical merits of the model, we refer to Taylor (1994), Ghysels et al. (1996) and Shephard (1996). The focus of this chapter are empirical estimation problems.
Despite the apparent simplicity of the specification, the basic SV model is not easy estimate, as no closed-form expression of the likelihood is available. Melino and Turnbull (1990) pioneered General Method of Moments based estimation of SV models. More elaborate method-of-moment procedures (Gallant et al., 1997; Andersen et al., 1999) improve upon the estimation efficiency, but they are generally outperformed by Markov Chain Monte Carlo (MCMC) estimation (Jacquier et al., 1994; Kim et al., 1998). Simulated Maximum Likelihood (Danielsson, 1994; Shephard and Pitt, 1997; Durbin and Koopman, 1997; Sandmann and Koopman, 1998; Liesenfeld and Richard, 2003) provides an attractive alternative that can be as efficient as MCMC while being easier to implement.

The model specification (7.1) can be converted directly into a linear state space model via a logarithmic transformation of the observations. The assumed Gaussianity of the disturbances imply that in the transformed linear model the observation disturbances follow log-Normal distributions. As noted in chapter 2, the standard Kalman filter provides minimum MSE estimates of the state (the log-volatility process) within the class of linear estimators. Maximizing the Gaussian likelihood in the linear model yields a quasi-maximum likelihood estimate of the parameters, see Harvey et al. (1994).

Model (7.1) can also be regarded as a non-linear state space model with multiplicative disturbances in the observation equations. It can be cast in the additive form by putting the disturbance $\epsilon_t$ in the state vector and specifying a non-linear observation equation with trivial disturbances. Alternatively, the non-linear estimation algorithms in chapter 2 can be adapted directly to account for the multiplicative disturbance factor, see Haykin (2001, Chap. 7). Unfortunately neither the extended nor the unscented Kalman filter is capable of dealing with the non-linearities in the SV model. Zoeter et al. (2004) provide a technical discussion of the shortcomings of these analytical approximation techniques for this problem, and develop a modified unscented filter to estimate SV models. As the modifications are not simple, this negates a major part of the attraction of using simple approximating filters. Quadrature based numerical integration is another approach to handle non-linearities. For SV models these have been developed by Fridman and Harris (1998), Bartolucci and Luca (2001), and in combination with Gaussian sum filters by Kawakatsu (2007).

In this chapter, we take a simulation based approach to estimating SV models. For similar reasons that the basic extended and unscented methods do not provide adequate estimates, importance samplers and particle filters based on these approximations are ill-suited to provide proposal densities for SV models. The simulation methods require approximations of the probability densities implied by the model, rather than of the functional form. We focus exclusively on importance sampling here, using algorithms developed in Sandmann and Koopman (1998) and Liesenfeld and Richard (2003). For
a particle filtering approach, we refer to Polson et al. (2003).

## 7.3 Likelihood estimation by importance sampling

The likelihood of the basic SV model (7.1) for observations \( \{y_1, \ldots, y_n\} \) is given by

\[
L(\phi_0, \phi_1, \sigma) = \int \cdots \int p(y_1, \ldots, y_n, h_1, \ldots, h_n) \, dh_1 \cdots dh_n. \tag{7.2}
\]

To improve legibility, we suppress the dependency of the density \( p(\cdot) \) on the parameters \((\phi_0, \phi_1, \sigma)\) in the notation. For the SV model, analytical solutions to the integrals in (7.2) are not available.

Writing \( y = \{y_1, \ldots, y_n\} \) and \( h = \{h_1, \ldots, h_n\} \), we see that the the likelihood

\[
\int p(y, h) \, dh = \int p(y|h) p(h) \, dh \tag{7.3}
\]

is the expectation of the conditional density \( p(y|h) \) under the distribution of \( h \). In principle, it can be estimated by generating \( N \) replications \( h^{(i)} \) from the unconditional density \( p(h) \) and computing the average \( N^{-1} \sum_{i=1}^{N} p(y|h^{(i)}) \). Since the SV model specifies \( p(y|h) \) as a Gaussian density with known mean and variance, and the dynamic property of \( h \) is implied by a standard AR(1) process, these calculations are straightforward.

The problem with such a naive simulation approach is that most replications \( h^{(i)} \) do not resemble the process \( h \) under which the observed \( y \) was obtained. For series of realistic length, \( N \) needs to be overwhelmingly large to estimate the likelihood with some accuracy. Importance sampling is an effective method to overcome this problem.

In the case of the SV model (7.1), we rewrite the likelihood as

\[
\int p(y|h) p(h) \, dh = \int \frac{p(y|h) p(h)}{g(h|y)} g(h|y) \, dh \tag{7.4}
\]

and estimate it by

\[
\hat{L}(\phi_0, \phi_1, \sigma) = \frac{1}{N} \sum_{i=1}^{N} \frac{p(y|h^{(i)}) p(h^{(i)})}{g(h^{(i)}|y)}, \tag{7.5}
\]

where simulated values for the processes \( h^{(i)} \) are sampled from the importance density \( g(h|y) \). The quality of the estimate depends on the choice of \( g(h|y) \), which should provide a good approximation to \( p(y,h) \) apart from a constant. In the next section, we describe two choices of the importance density, and provide details of the algorithms for generating importance samples.
7.4 Obtaining importance densities

Method 1  The first method we describe is the importance sampling procedure developed in Shephard and Pitt (1997) and Durbin and Koopman (1997). It is based on a linear Gaussian state space approximation to the original model. Equation (7.1b) provides the transition equation of the state variable $h_t$, while the observation equation is given by

$$
y_t = h_t + a_t + b_t u_t.
$$

(7.6)

The standard Gaussian noise process $\{u_t\}$ is assumed to be uncorrelated with the state disturbance process $\{\eta_t\}$. The conditional density $g(h|y)$ of the approximating model serves as the importance density.

The location and scale parameters $a_t$ and $b_t$, respectively, are chosen to provide a good match between the original and the approximating model. This is done by equalising the first and second derivatives of $\log g(y_t|h_t)$ with respect to $h_t$ for both the original and the approximating model. The parameters $a_t$ and $b_t$ can be solved from the two equations for given values of $h$ and $y$. The process $h_t$ is not observed, but can be estimated using the Kalman smoother. At the start, neither $a_t$, $b_t$, nor $h_t$ are known. The algorithm is started with some trial $h_t$ and alternates between solving $a_t$, $b_t$ and estimating $h_t$ using the Kalman smoother until either converges.

The converged model is a basic linear Gaussian state space model. The best estimates of the state mean and variance given $y$ are produced by the Kalman smoother. Samples from the importance density $g(h|y)$ can subsequently be obtained via simulation smoothing techniques as described in

In summary, the algorithm to calculate the likelihood of $y$ given parameters $(\phi_0, \phi_1, \sigma)$ proceeds as follows:

1. Choose a trial $h$.

2. Solve $a_t$ and $b_t$ by equalizing derivatives of the log-density of the original and approximating model.

3. Estimate $h_t$ from the linear state space model defined by (7.6), (7.1b), (7.1c) with the Kalman smoother.

4. Repeat 2. and 3. until convergence.

5. Estimate the likelihood according to equation (7.5), where samples $h(t)$ are drawn from $g(h(t)|y)$ with a simulation smoother, and the density $g(h(t)|y)$ is calculated by the Kalman smoother.

More details can be found in appendix B.
The second importance sampler that we investigate is developed in Richard and Zhang (2000) and Liesenfeld and Richard (2003). It is based on a particular factorization of the importance density. The importance density \( g(h|y) \) is constructed as

\[
g(h|y) = \prod_{t=1}^{n} g(h_t|h_{t-1}, y) = \prod_{t=1}^{n} C_t \times \exp(c_t h_t + d_t h_t^2) \times p(h_t|h_{t-1}), \tag{7.7}
\]

where \( C_t \) is chosen such that \( g(h_t|h_{t-1}, y) \) is Gaussian, and the coefficients \( c_1, \ldots, c_n \) and \( d_1, \ldots, d_n \) are chosen to ensure that \( g(h|y) \) resembles \( p(y, h) \) for the given observations \( y \). The Gaussianity restrictions imply the values of \( C_t \) when \( c_t, d_t \) are given, as is shown in appendix C. For a set of \( N \) simulated draws of \( h^{(i)} \) drawn from \( g(h|y) \), finding appropriate coefficients \( c_t, d_t \) to provide good match between \( g(h^{(i)}|y) \) and \( p(y, h^{(i)}) \) is a non-linear high dimensional optimization problem, which can be simplified as follows. From (7.5), the importance sampling likelihood estimate is given by

\[
\hat{L}(\phi_0, \phi_1, \sigma) = \frac{1}{N} \sum_{i=1}^{N} \prod_{t=1}^{n} \frac{p(y_t|h_t^{(i)})p(h_t^{(i)}|h_{t-1}^{(i)})}{C_t \exp(c_t h_t^{(i)} + d_t h_t^{(i)2})p(h_t|h_{t-1})} = \frac{1}{N} \sum_{i=1}^{N} \prod_{t=1}^{n} \frac{p(y_t|h_t^{(i)})}{C_t \exp(c_t h_t^{(i)} + d_t h_t^{(i)2})}. \tag{7.8}
\]

Expanding the product from \( t = 1, \ldots, n \) and observing that \( C_t \) depends on \( h_{t-1} \) but not on \( h_t \), it can be seen that a good match can be obtained by choosing \( c_t \) and \( d_t \) such that \( p(y_t|h_1^{(i)}) \) resembles \( C_1 C_2 \exp(c_1 h_1^{(i)} + d_1 h_1^{(i)2}) \), and \( p(y_t|h_t^{(i)}) \) resembles \( C_{t+1} \exp(c_t h_t^{(i)} + d_t h_t^{(i)2}) \) for \( t = 2, \ldots, n \), where we define \( C_{n+1} = 1 \). Operationally, this amounts to calculating least squares coefficients \( c_t, d_t \) of the regression equations

\[
\log p(y_t|h_t^{(i)}) - \log C_{t+1} = \text{constant} + c_t h_t^{(i)} + d_t h_t^{(i)2}, \quad i = 1, \ldots, N, \tag{7.9}
\]

where \( h_t \) and \( h_t^2 \) are the explanatory variables. Since \( C_{t+1} \) depends on \( c_{t+1} \) and \( d_{t+1} \), the coefficients are calculated recursively, proceeding from \( t = n, n-1, \ldots, 1 \).

The parameters \( c_t \) and \( d_t \) of the approximating model are calculated for given \( y, (\phi_0, \phi_1, \sigma) \) and \( h^{(i)}, i = 1, \ldots, N \). We can start the algorithm by drawing \( h^{(i)} \) unconditionally according to (7.1b), (7.1c). With the estimated \( c_t, d_t \), a new set of \( h^{(i)} \) is drawn from the Gaussian density \( g(h|y) \). This is iterated until either \( h^{(i)} \) or \( c_t, d_t \) converge.

Summarizing, the algorithm to calculate the likelihood of \( y \) given parameters \((\phi_0, \phi_1, \sigma)\) proceeds as follows:

1. Draw samples \( h^{(i)} \) according to (7.1b), (7.1c).
2. Solve \( c_t \) and \( d_t \) by performing \( n \) regressions as in equation (7.9), working backwards from \( t = n \) until \( t = 1 \), with \( C_{n+1} = 1 \).
3. Draw samples $h^{(i)}$ from the density in equation (7.7).

4. Repeat 2. and 3. until convergence.

5. Estimate the likelihood according to equation (7.8), where samples $h^{(i)}$ are drawn as in step 3.

The explicit formula for $C_t$ is given in appendix C.

7.5 Simulation results

We implemented the two methods described in the previous section, which we will refer to as IS1 and IS2 respectively, in the Ox programming language version 3.30 (Doornik, 2002). The Kalman filter and simulation smoother used in IS1 were calculated with the use of SsfPack 2.3 (Koopman et al., 1999).

To compare the performance of the presented algorithms, we carry out a simulation study in which data was generated according to the true model. We draw $1,000$ series of $n = 1,000$ observations, with parameter vector $(\phi_0, \phi_1, \sigma) = (0.0, 0.97, 0.2)$. Such parameter values are fairly representative of what is found in empirical work with financial assets and currency exchange rate returns. For each generated series, the likelihood calculated by IS1 and IS2\(^1\) were maximised with the BFGS algorithm, starting at the true parameters. To enhance the stability of the numerical optimization procedure, the likelihood was calculated as a function of the parameters $(\phi'_0, \phi'_1, \sigma') = (\phi_0, \log(\phi_1/(1-\phi_1)), \log \sigma)$. These transformations result in distributions of the estimators that are considerably less skewed than those of the original parameters. The transformed values of the true parameter are $(\phi'_0, \phi'_1, \sigma') = (0.0, 3.48, -1.61)$.

In table 7.1 and figure 1 we present the simulation results. The first two moments do not differ very much. The IS2 estimate of $\phi'_1$ appears to be somewhat less biased\(^2\) than IS1, which can be ascribed to a number of large positive outliers. The IS1 estimates are more symmetric for all three parameters than the IS2 estimates. The kurtosis is also consistently lower. From the graphs it is especially evident that for IS2, the estimates of $\phi'_1$ are especially prone to exhibiting large outliers.

\(^1\)We used $N = 200$ and $N = 50$ for IS1 and IS2 respectively. Both values are higher that the recommended values in the original implementations. With these values, IS1 is slightly faster in our current implementation.

\(^2\)The bias may be corrected to some degree, see Durbin and Koopman (1997).
Table 7.1: parameter estimates

<table>
<thead>
<tr>
<th></th>
<th>IS1</th>
<th></th>
<th>IS2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>true value</td>
<td>0.00</td>
<td>3.48</td>
<td>-1.61</td>
<td>0.00</td>
</tr>
<tr>
<td>mean</td>
<td>-0.00</td>
<td>3.17</td>
<td>-1.57</td>
<td>-0.00</td>
</tr>
<tr>
<td>st. deviation</td>
<td>0.01</td>
<td>0.52</td>
<td>0.24</td>
<td>0.01</td>
</tr>
<tr>
<td>skewness</td>
<td>0.04</td>
<td>-0.35</td>
<td>-0.28</td>
<td>-0.56</td>
</tr>
<tr>
<td>kurtosis</td>
<td>2.97</td>
<td>3.74</td>
<td>4.79</td>
<td>6.48</td>
</tr>
</tbody>
</table>

Figure 7.1: Empirical densities of parameter estimates
Table 7.2: Likelihood based tests for the null of the existence of the variance of the importance sampler. Score and t-tests are both standard normal, with the test being rejected for large positive values. LR test has a $0.5 (\chi_0^2 + \chi_1^2)$ null. It is rejected for large values. The 95% critical values are 1.65, 1.65 and 2.69, respectively.

### 7.6 Importance sampling diagnostics

The importance of testing the validity of the finite second moment assumption in importance sampling has been noted in section 3.2.3. In this section we show some results of the testing procedure described in that chapter.

Figure 7.2 show simple diagnostic results for the IS1 and IS2 methods computed with $N = 100,000$ draws. It is apparent that IS1 has some very large outliers. The weights for IS2 are better behaved, but the size of the outliers is still quite large, compared to that of typical values which appear in the histogram. The recursive standard deviations of the weights, displayed in the bottom panels were calculated from $N = 1,000$ onwards, which is much higher than usual.\(^3\) The variance of the weights is clearly not very stable, even with such a high number of draws.

The calculated values of three test statistics are reported in table 7.2. The estimated value of $\xi$ for IS2 was smaller than $1/2$, implying that we cannot reject the existence of the variance. For IS1, the null is easily rejected. It appears that IS2 is more reliable for estimating the likelihood than IS1, whereas earlier we found that the actual parameter estimates provided by IS1 were slightly more reliable than those calculated from IS2.

### 7.7 Stochastic volatility with t disturbances

The sample kurtosis in empirical return series is often larger than $3 \exp(\sigma^2/(1 - \phi_1^2))$, which is implied by the the Gaussian SV specification, see Shephard (1996). A straightforward modification of the basic model that captures this effect is the assumption of a leptokurtic distribution for the disturbance $\epsilon_t$. The Student-t distribution is a simple and attractive choice, as the degrees of freedom $\nu$ parameterises the tail fatness. For high values of $\nu$ the model practically nests the basic Gaussian model. We will refer to


<table>
<thead>
<tr>
<th></th>
<th>IS1</th>
<th>IS2</th>
</tr>
</thead>
<tbody>
<tr>
<td>score</td>
<td>4.63</td>
<td>-3.83</td>
</tr>
<tr>
<td>t-value</td>
<td>8.27</td>
<td>-7.39</td>
</tr>
<tr>
<td>LR</td>
<td>86.4</td>
<td>-</td>
</tr>
</tbody>
</table>
model (7.1) in which $\epsilon_t$ is a sequence of independent $t$ distributed variables scaled to unit variance, as the SV$t$ model. This model also was analysed in Harvey et al. (1994), Sandmann and Koopman (1998) and Liesenfeld and Richard (2003).

Both importance sampling methods can be easily adapted to estimate the SV$t$ model. For IS1, the procedure described in the summary in section 3 remains the same, with the Gaussian density $p(y|h)$ replaced by a scaled $t$ density. This means that in step 2 a different set of equations must be solved, and in step 5 the calculation of $\hat{L}$ is performed with a $t$ density in the numerator of equation (7.5). Note that the approximating state space model (and hence the importance density) remains Gaussian, only the way the approximation is obtained through $a_t$ and $b_t$ is altered. Appendix B provides the relevant details.

The adaptation of IS2 is similar to that of IS1. In step 2 of the summary at the end of section 3, the regression is performed with a $t$ density for $p(y_t|h_t)$ in equation (5). Again, the importance density remains Gaussian by construction; it adapts to the altered true model through $c_t$ and $d_t$. In the calculation of $\hat{L}$ in step 5 the Gaussian density for $p(y|h)$ is replaced by a $t$ density.

We maximised the likelihood functions with respect to the transformed degrees of freedom parameter $\nu' = \log(\nu - 2)$ rather than $\nu$. We assume that $\epsilon_t$ has a finite variance, which implies that $\nu > 2$. The estimated parameters are presented in table 7.3.

Figure 7.2: Importance weights diagnostics, from top to bottom: 100 largest weights, histogram excluding 100 largest weights, recursive standard deviation.
### 7.7. Stochastic Volatility with T Disturbances

<table>
<thead>
<tr>
<th></th>
<th>IS1</th>
<th>IS2</th>
</tr>
</thead>
<tbody>
<tr>
<td>true value</td>
<td>1.79 0.00 3.48 -1.61</td>
<td>1.79 0.00 3.48 -1.61</td>
</tr>
<tr>
<td>mean</td>
<td>1.94 -0.00 3.35 -1.61</td>
<td>2.31 -0.00 3.29 -1.55</td>
</tr>
<tr>
<td>st. deviation</td>
<td>0.62 0.01 0.45 0.22</td>
<td>0.68 0.01 0.56 0.27</td>
</tr>
<tr>
<td>skewness</td>
<td>5.02 0.01 0.14 -0.43</td>
<td>5.87 -1.42 0.32 -0.41</td>
</tr>
<tr>
<td>kurtosis</td>
<td>55.1 3.01 3.00 3.77</td>
<td>63.0 12.4 5.70 3.99</td>
</tr>
</tbody>
</table>

**Table 7.3: SVt parameter estimates**

![Empirical densities of SVt parameter estimates](image)

#### Figure 7.3: Empirical densities of SVt parameter estimates

<table>
<thead>
<tr>
<th></th>
<th>IS1</th>
<th>IS2</th>
</tr>
</thead>
<tbody>
<tr>
<td>score</td>
<td>-2.40</td>
<td>-1.24</td>
</tr>
<tr>
<td>t-value</td>
<td>-4.93</td>
<td>-2.43</td>
</tr>
<tr>
<td>LR</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

**Table 7.4: Likelihood based tests for the existence of the variance of the importance sampler in the SVt model.** Score and t-tests are both standard normal, with the test being rejected for large positive values. $\hat{\xi}_t$ was negative for both methods, indicating finite variance.
and figure 7.3. Compared to the estimates for the model with Gaussian disturbances, the corresponding parameter estimates change marginally. For the IS2 method the large outliers have disappeared. The estimates of the new parameter $\nu'$ are less precise than the other three parameters. With both methods we encountered some great outliers which are largely responsible for the upwards bias, significant skewness and large kurtosis of $\hat{\nu}'$. The estimates for IS2 are slightly worse than for IS1.

In figure 7.4 we show the diagnostic graphs for the importance weights as documented in section 5. Compared to figure 7.2, we observe that the weights for IS1 are much more stable. IS2 only shows a slight improvement. For the SVt model the weights from IS1 are more stable than from IS2. Fitting a Pareto distribution to the tail weights, the estimated value of $\xi$ is negative. Thus, we have no evidence that the variation of the importance weights is not bounded. This is confirmed by the likelihood tests presented in table 7.4.

### 7.8 Conclusion

Stochastic volatility models can be estimated by classical importance sampling methods as is shown by various contributions in the econometrics literature. Two distinct implementations are investigated in detail for the basic model with noise processes from
7.8. **CONCLUSION**

Gaussian and heavy-tailed t densities. In particular a simulation study is employed that generates data from the true model and estimates the parameters of the model given the simulated data. By repeating these step many times, an assessment can be given about the accuracy of both implementations. It is found that for the model with Gaussian disturbances, the implementation IS1 gives more stable parameter estimates than IS2. Further, the importance samplers themselves can be diagnosed using descriptive statistics and graphics and using more formal test statistics based on extreme-value distributions. Based on these diagnostics, IS2 can be regarded as a more reliable importance sampler compared to IS1.

In the case of t distributed disturbances, we found no problems with either method. Both the importance sampling procedure and the parameter estimates gave stable results. As the SVt model is usually seen as a better match for market data than the Gaussian SV model (Sandmann and Koopman, 1998; Liesenfeld and Richard, 2003), we would recommend to use this as the base model in empirical work. A slight drawback is the limited accuracy of the estimate of the tail fatness, although this phenomenon is not specific to the SV model.

**Appendix A: Probability densities**

The Gaussian log-densities used in the basic SV model are given by

\[
\log p(y|h) = -\frac{T}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{T} (y_t^2/e^{h_t} + h_t),
\]

\[\log p(h) = \log p(h_1) - \frac{T-1}{2} \log 2\pi \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^{T-1} (h_{t+1} - \phi_0 - \phi_1 h_t)^2, \tag{7.10b}\]

\[
\log p(h_1) = -\frac{1}{2} \log \frac{2\pi \sigma^2}{1 - \phi_1^2} - \frac{1}{2} \left( h_1 - \frac{\phi_0}{1 - \phi_1} \right)^2 / \frac{\sigma^2}{1 - \phi_1^2}. \tag{7.10c}
\]

In the SVt specification, \(\epsilon_t\) in equation (7.1a) is a scaled Student-t variable with variance 1 and \(\nu\) degrees of freedom, rather than a standard Gaussian variable. This implies that the log-density in (7.10) becomes

\[
\log p(y|h) = T \log \frac{\Gamma \left( \frac{\nu}{2} + \frac{1}{2} \right)}{\Gamma \left( \frac{\nu}{2} \right)} - \frac{T}{2} \log(\nu - 2) - \frac{1}{2} \sum_{t=1}^{T} h_t
\]

\[- \frac{1}{2} \sum_{t=1}^{T} (\nu + 1) \log \left( 1 + \frac{y_t^2 \exp(-h_t)}{\nu - 2} \right). \tag{7.11}\]
Appendix B: IS1

The observation equation (7.6) implies the conditional observation log-density
\[ g_t = \log g(y_t|h_t) = -\frac{1}{2} \log 2\pi b_t^2 - \frac{1}{2} (y_t - h_t - a_t)^2 / b_t^2. \] (7.12)

Its first two derivatives with respect to \( h_t \) are
\[ \dot{g}_t = (y_t - h_t - a_t) / b_t^2, \quad \ddot{g}_t = -1 / b_t^2. \] (7.13)

From (12a), the derivatives of the log-density \( p_t = \log p(y_t|h_t) \) are
\[ \dot{p}_t = \frac{1}{2} y_t^2 / e^{h_t} - \frac{1}{2}, \quad \ddot{p}_t = -\frac{1}{2} y_t^2 / e^{h_t}. \] (7.14)

From equating the derivatives, we obtain the solutions
\[ b_t^2 = 2 e^{h_t} / y_t^2, \quad a_t = \frac{1}{2} b_t^2 + y_t - h_t - 1. \] (7.15)

In the SVt specification, \( p_t \) is given by (7.11), which has derivatives
\[ \dot{p}_t = \frac{\nu + 1}{2} \left( 1 + (\nu - 2) e^{h_t} / y_t^2 \right)^{-1} - \frac{1}{2}, \quad \ddot{p}_t = -\frac{1}{2} \frac{(\nu + 1)(\nu - 2)}{2 y_t^2 / e^{h_t}} \left( 1 + (\nu - 2) e^{h_t} / y_t^2 \right)^{-2}. \] (7.16)

Equating these to (7.13), we obtain
\[ b_t^2 = \frac{2}{\nu + 1} \left( \frac{y_t^2}{(\nu - 2) e^{h_t}} + 2 \right) + \frac{(\nu - 2) e^{h_t}}{y_t^2}, \quad a_t = y_t - h_t - b_t^2 \dot{p}_t. \] (7.17)

For given observations \( y \) and parameters \( (\phi_0, \phi_1, \sigma) \), the solution of \( a_t \) and \( b_t \) also require the latent \( h \). In practice, this can be obtained by starting with a trial \( h \) and calculating the approximating model. Given the approximating model, a new estimate of \( h \) is obtained from the Kalman smoother. This is iterated until either the estimated \( h \) or \( a_t \) and \( b_t \) converge.

Writing the conditional density of the approximating model as \( g(y|h) \), we can use a simulation smoother to obtain replications \( h^{(i)} \) drawn from \( g(h|y) \) and to calculate the importance density \( g(h|y) \). Note that the approximating marginal density \( p(h) \) is the same as in the original model. Substituting \( g(h|y) = g(y|h)p(h)/g(y) \) in the likelihood equation yields
\[
L(\phi_0, \phi_1, \sigma) = \int \frac{p(y|h)p(h)}{g(h|y)} g(h|y) \, dh = g(y) \int \frac{p(y|h)p(h)}{g(y|h)p(h)} g(h|y) \, dh = g(y) \int \frac{p(y|h)}{g(y|h)} g(h|y) \, dh.
\] (7.18)
Thus, the likelihood can be estimated by multiplying the Gaussian likelihood of the approximating model \( g(y) \) by the correction factor

\[
\frac{1}{N} \sum_{i=1}^{N} \frac{p(y|h^{(i)})}{g(y|h^{(i)})}
\]

(7.19)

obtained from simulation.

Appendix C: IS2

If \( h_t \) is Normally distributed with mean \( M_t \) and variance \( S_t^2 \) its log-density is given by

\[
-\frac{1}{2} \log 2\pi S_t^2 - \frac{1}{2} \left( \frac{h_t - M_t}{S_t} \right)^2 = -\frac{1}{2} \log 2\pi S_t^2 - \frac{1}{2S_t^2} h_t^2 + \frac{M_t}{S_t^2} h_t - \frac{M_t^2}{2S_t^2}.
\]

(7.20)

From (7.7) and (12b) the function \( \log g(h_t|h_{t-1}, y) \) is given by

\[
\log g(h_t|h_{t-1}, y) = \log C_t + c_t h_t + d_t h_t^2 + \log p(h_t|h_{t-1})
\]

\[
= \log C_t + c_t h_t + d_t h_t^2 - \frac{1}{2} \log 2\pi \sigma^2 - \frac{1}{2} \left( \frac{h_t - \phi_0 - \phi_1 h_{t-1}}{\sigma} \right)^2
\]

\[
= \log C_t - \frac{1}{2} \log 2\pi \sigma^2 - \frac{1}{2} \left( \frac{1}{\sigma^2} - 2d_t \right) h_t^2 + \left( \frac{\phi_0 + \phi_1 h_{t-1}}{\sigma} + c_t \right) h_t
\]

\[- \frac{(\phi_0 + \phi_1 h_{t-1})^2}{2\sigma^2}.
\]

(7.21)

Equating this expression with (7.20) it follows that \( \log g(h_t|h_{t-1}, y) \) is the log-density of a Normal variable with variance

\[
S_t^2 = \frac{\sigma^2}{1 - 2d_t \sigma^2}
\]

(7.22)

and mean

\[
M_t = \left( \frac{\phi_0 + \phi_1 h_{t-1}}{\sigma^2} + c_t \right) S_t^2
\]

(7.23)

if we choose

\[
\log C_t = -\frac{M_t^2}{2S_t^2} + \frac{(\phi_0 + \phi_1 h_{t-1})^2}{2\sigma^2} - \frac{1}{2} \log \frac{S_t^2}{\sigma^2}.
\]

(7.24)
Chapter 8

Conclusions

8.1 General remarks

The usefulness of linear state space modelling has been demonstrated in many economic applications in the past few decades. Presumably, the far less restrictive class of non-linear state space models is applicable to a wider range of problems. However, general state space models are considerably more challenging to use than linear models. In the non-linear setting, the researcher is confronted with a multitude of options, in both model formulation and estimation technology. In either case, guidance from theory is rather limited. On the estimation front, techniques are far less precise compared to the linear state space setting. The introduction of the non-linear chapter in the classic textbook *Optimal Filtering* (Anderson and Moore, 1979) warns that “we must work hard to achieve even a little”.

In this dissertation, an attempt has been made to demonstrate both practical estimation techniques and algorithms, as well as economically relevant models. Both for estimation and model formulation, we depart from an established linear framework, and often do not stray too far. The results can hopefully encourage researchers operating near the boundaries in the linear state space settings to take a few more steps. Much new research on estimating non-linear state space has been conducted since the publication of *Optimal Filtering* by Anderson and Moore (1979), and while we still need to work hard, we can now often achieve considerably more than we could in the past.

8.2 Methods

Non-linear filtering techniques can generally be divided in simulation, or Monte-Carlo based algorithms, and non-simulation based algorithms. In chapter 2, we examine two non-simulation based algorithms in detail, namely the extended and the unscented
Kalman filter. The extended Kalman filter is the oldest and most established non-linear state space filtering algorithm. It is derived as the application of the standard Kalman filter to a first order approximation of the non-linear model. In this chapter, the extended filter is also provided for models with diffuse initial conditions.

The unscented Kalman filter is a more recent non-linear filter which is gaining acceptance in engineering disciplines. Using an innovative transformation, it utilises a higher order Taylor approximation without the need to calculate higher order derivatives.

Chapter 3 describes Monte Carlo estimation. Although naive Monte Carlo estimators are readily derived from the Bayesian recursive solution to the filtering problem, any practical application requires some variance reduction techniques. This chapter concentrates on importance sampling, both in a global form, where an approximating model is built for the entire state process before the simulation commences, as well as in a sequential form in which approximating model are obtained using simulations in each time step. The latter approach is usually termed particle filtering.

### 8.3 Applications

The next four chapters demonstrate economic applications of general state space models. The first three are extensions of the well-known unobserved components (UC) models, in which time series are decomposed in stochastic processes which represent trends, cycles, seasonal effects, noise and possibly other sources of variation.

In chapter 4 logistic and smooth spline functions are used to permit variation in the evolution of the parameters of the cycle component. This is strictly speaking a linear state space model in the sense that the transition and observation processes are both linear. The parameters are transformed as a non-linear deterministic functions of time, which allows more flexibility compared to the basic specification. The specification appears to be suitable for U.S. macro economic times series, and is relatively easy to estimate. However, this comes at the cost of having to use many parameters.

Chapter 5 develops a non-linear cycle specification in which the length of the period or the frequency depends on the steepness of the cycle. This can capture asymmetry in the cycle, a phenomenon often observed in macro-economic time series, in which expansions and contractions occur at different velocities. The asymmetry parameter is estimated using an importance sampling based likelihood. In the U.S. macro-economic time series we estimate that GDP and investment tend to fall significantly faster than that the rise, while unemployment increases faster than it decreases.

In chapter 6, the seasonal component in a UC model is linked to both the trend and the cycle, such that the degree of seasonal variation may depend on the other components. We show that this specification can be used to model increasing seasonal
variation in cases where a logarithmic transformation is not appropriate, and that it captures cycle dependent seasonality. In U.S unemployment it is found that seasonal effects are more pronounced during expansions than during recessions, and that in dwellings productions seasonality is attenuated in the cycle and has decreased overall together with the rise in the trend.

Finally, in chapter 7 two importance sampling based estimators are compared in their effectiveness to estimate stochastic volatility models. The performance between the methods which arrive at model approximations from rather different paths are comparable, and both methods can be used effectively to estimate stochastic volatility model with Normal and Student t disturbances.
Bibliography


In de state space methodiek worden indirect waargenomen dynamische verschijnselen gemodelleerd met een wiskundige beschrijving van de dynamiek en van de wijze van waarneming. Oorspronkelijk werden dergelijke modellen toegepast in de ontwikkeling van ruimtevaart technologie om de locatie van objecten zoals raketten te bepalen uit indirecte en niet geheel betrouwbare meetgegevens. Door de algemeenheid van de modellen en de rekenmethoden om de indirect waargenomen toestand-variabelen te schatten, werden state space modellen in toenemende mate gebruikt in andere disciplines. Globaal kan een state space model worden samengevat als een model van een verschijnsel dat verandert in de tijd, en dat niet direct gemeten of waargenomen wordt. Het is niet verwonderlijk dat in de economische wetenschap veel problemen binnen dit kader te plaatsen zijn. Belangrijke voorbeelden zijn het schatten van de conjunctuurfase, seizoensinvloeden op economische activiteit, het algemene prijsspeil, de bewegelijkheid van prijzen en dynamische parameters (welke in traditionele regressie-analyses doorgaans constant worden verondersteld).

Het centrale algoritme binnen de state space methodiek is een verzameling vergelijkingen welke bekend staan onder de naam Kalman filter. Met dit algoritme kan de statistisch optimale schatting van de indirect waargenomen variabelen worden berekend. Rekentechnisch is het Kalman filter efficiënt door de recursieve opbouw; nieuwe waarnemingen worden telkens gecombineerd met de tot dusverre berekende optimale schatting om tot een nieuwe schatting te komen. In de klasse van lineaire state space modellen met Normaal verdeelde variabelen geeft het Kalman filter de optimale schatting van de toestand-variabelen (in de zin van de kleinste verwachte kwadratische fout). Als de dynamische vergelijking of de waarnemingsvergelijking in het state space model echter niet lineair zijn, is de optimale schatting niet eenvoudig te berekenen. De oplossing bestaat uit integraal-vergelijkingen van vaak hoge dimensie en zonder oplossingen in gesloten vorm. Voor praktische toepassingen is in de regel een benadering van de exacte oplossing noodzakelijk.

In het introducte hoofdstuk wordt het algemene niet-lineaire state space model besproken, en ingegaan op het schattingsprobleem. De precieze wiskundige formuleringen en de algemene oplossing van optimale schattingen zijn hierin te vinden.
Benaderingen van de optimale schatter kunnen ingedeeld worden in deterministische en simulatie gebaseerde methoden. De deterministische algoritmen genaamd het extended Kalman filter en het unscented Kalman filter komen uitgebreid aan bod in hoofdstuk 2. Parameter schattingen, voorspellingen en schattingen gebaseerd op complete databestanden (het zogenaamde smoothen) komen aan bod. Voor het extended Kalman filter wordt getoond hoe het startwaarde probleem (waarbij in de eerste stap van de recursieve vergelijkingen vaak geen schatting van de toestand-variabelen beschikbaar is) opgelost kan worden op vergelijkbare wijze als in lineaire modellen. Verder wordt een vergelijking gegeven tussen smoothing gebaseerd op extended en unscented Kalman filters. In hoofdstuk 3 worden simulatietechnieken besproken die toegepast kunnen worden in niet-lineaire state space modellen. Dergelijke methoden zijn bewerkelijker dan de deterministische methoden uit hoofdstuk 2, en vereisen veel meer rekentijd. De schattingen zijn echter doorgaans nauwekeuriger, daar ze de theoretisch optimale schatter willekeurig dicht kunnen benaderen. Uitgebreide aandacht wordt besteed aan importance sampling methoden, welke uitgaan van eenvoudige benaderende modellen als basis voor de simulaties. Importance sampling simulaties kunnen globaal over alle data worden uitgevoerd, of sequentieel in de tijd, in welk geval ze vaak particle filters worden genoemd.

In hoofdstukken 4 tot en met 7 zijn enkele toepassingen van niet-lineaire state space modellen te vinden. In een veel gebruikte model-type genaamd unobserved components (UC) modellen, worden datareeksen verondersteld te zijn opgebouwd uit verschillende componenten die niet afzonderlijk zijn waar te nemen. Voor economische analyse zijn de componenten bijvoorbeeld een trend, welke relatief langzaam veranderd, een seizoens-effect, welke zich periodiek herhaald met een hoge frequentie, een cyclische component, welke een middel-lange periode kent (ook bekend als een conjunctuurgolf), en een onregelmatige verstoring die we niet verder trachten te modelleren of voorspellen. De simulaties kunnen globaal over alle data worden uitgevoerd, of sequentieel in de tijd, in welk geval ze vaak particle filters worden genoemd.

Hoofdstuk 5 maakt het seizoen-component afhankelijk van de trend en van het cyclische component. De trend-afhankelijkheid biedt een alternatieve manier om de
grootte van seizoenscorrecties te relateren aan de onderliggende grootheid, wanneer een eenvoudige data-transformatie ongewenst of niet adequaat lijkt. Interactie van seizoens-effecten met het cyclische component maakt het mogelijk om de invloed van het seizoen in verschillende fasen van de conjunctuur te varieren. Toepassingen van deze modellen worden getoond met tijdreeksen van aantallen reizen van en naar het buitenland, werkloosheid en huisproductie.

Hoofdstuk 7 behandelt een simulatiestudie met stochastische volatiliteit. Hierbij worden tijdreeksen gegenereerd waarin langere perioden met gemiddeld grotere en kleinere bewegingen voorkomen. Om het variantie-process (welke de beweeglijkheid van een datareeks weerspiegeld) te achterhalen uit de datareeks kan een niet-lineair state space model worden gebruikt. Bekende deterministische methoden blijken ontevreden. Met simulatie-gebaseerde schattingmethoden is het variantie-process echter goed te achterhalen.

Hoofdstuk 8 sluit het proefschrift af met algemene conclusies.
SAMENVATTING (SUMMARY IN DUTCH)
Acknowledgements

Conducting PhD research has often been compared to a lonely journey. Nevertheless, when a dissertation is finalised, many thanks are always due to the interesting fellow travellers who have shared some parts of the journey, short or long. My first expression of gratitude must go to my supervisor Siem Jan Koopman. Without his boundless optimism, devotion and expertise in all areas of my research, I would not have reached the destination. I have great admiration and respect for the enthusiasm and energy that he brings to all the projects that he manages. I am also grateful to all members of the committee who have approved my dissertation for public defence; for their thorough reading, criticism and corrections to the manuscript.

I wish to thank especially Aart de Vos, who I always seem to meet at at critical junctions on my journey. Although I have never become a fully converted Bayesian, I have at least taken the lessons on Kalman filtering to heart. I also thank Marc Francke for supervision in applied econometric analysis and filtering before (and as it appears, also after) my PhD research.

I thank all my former colleagues at the VU econometrics department for making the place interesting to work. Special mention must go to the core time series group Marius, Charles, and later Drew, Michael and Jacques, and Brian and Suncica and all other fellow graduate students for the occasional lunch, conversation and day trips. I am grateful for having shared an office and various research and teaching duties with Suncica for many years. Her efforts to improve the working environment at both the TI and the VU have benefited many of us.

All people from my family have always been supportive of me, and I wish to thank them here. In particular, I thank my parents and my brother for their unconditional love and trust throughout all my life. I owe them more than I can express in words.

Finally I thank Irma for sharing everything with me. Our journeys have now become one, and I’m confident it will still be many years from now, when our student years have become memories of a distant past.
The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:


421. R.K. ANDADARI, *Local clusters in global value chains, a case study of wood furniture clusters in Central Java (Indonesia).*

422. V. KARTSEVA, *Designing Controls for Network Organizations: A Value-Based Approach.*


425. M. VAN DER VOORT, *Modelling Credit Derivatives.*


437. R. LORD, Efficient pricing algorithms for exotic derivatives.


439. Y.-Y. TSENG, Valuation of travel time reliability in passenger transport.

440. M.C. NON, Essays on Consumer Search and Interlocking Directorates.

441. M. DE HAAN, Family Background and Children’s Schooling Outcomes.

442. T. ZAVADIL, Dynamic Econometric Analysis of Insurance Markets with Imperfect Information.

443. I.A. MAZZA, Essays on endogenous economic policy.

444. R. HAIJEMA, Solving large structured Markov Decision Problems for perishable-inventory management and traffic control.


446. R. SEGERS, Advances in Monitoring the Economy.


448. L. PAN, Poverty, Risk and Insurance: Evidence from Ethiopia and Yemen.


454. S. VUJIĆ, Econometric Studies to the Economic and Social Factors of Crime.

455. F. HEUKELOM, Kahneman and Tversky and the Making of Behavioral Economics.


460. O.E. JONKEREN, *Adaptation to Climate Change in Inland Waterway Transport.*


462. J. NIEMCZYK, *Consequences and Detection of Invalid Exogeneity Conditions.*


464. M. KRAWCZYK, *Affect and risk in social interactions and individual decision-making.*


469. J.L.W. VAN KIPPERSLUIS, *Understanding Socioeconomic Differences in Health: An Economic Approach.*


