1 Introduction

1.1 Basic ideas of state space analysis

State space modelling provides a unified methodology for treating a wide range of problems in time series analysis. In this approach it is assumed that the development over time of the system under study is determined by an unobserved series of vectors $\alpha_1, \ldots, \alpha_n$, with which are associated a series of observations $y_1, \ldots, y_n$; the relation between the α_t's and the y_t's is specified by the state space model. The main purpose of state space analysis is to infer the relevant properties of the α_t's from a knowledge of the observations $y_1, \ldots, y_n$. Other purposes include forecasting, signal extraction and estimation of parameters. This book presents a systematic treatment of this approach to problems of time series analysis.

Our starting point when deciding the structure of the book was that we wanted to make the basic ideas of state space analysis easy to understand for readers with no previous knowledge of the approach. We felt that if we had begun the book by developing the theory step by step for a general state space model, the underlying ideas would be obscured by the complicated appearance of many of the formulae. We therefore decided instead to devote Chapter 2 of the book to a particularly simple example of a state space model, the local level model, and to develop as many as possible of the basic state space techniques for this model. Our hope is that this will enable readers new to the techniques to gain insights into the ideas behind state space methodology that will help them when working through the greater complexities of the treatment of the general case. With this purpose in mind, we introduce topics such as Kalman filtering, state smoothing, disturbance smoothing, simulation smoothing, missing observations, forecasting, initialisation, maximum likelihood estimation of parameters and diagnostic checking for the local level model. We present the results from both classical and Bayesian standpoints. We demonstrate how the basic theory that is needed for both cases can be developed from elementary results in regression theory.

1.2 Linear models

Before going on to develop the theory for the general model, we present a series of examples that show how the linear state space model relates to problems of practical interest. This is done in Chapter 3 where we begin by showing how structural time series models can be put into state space form. By structural time series models we mean models in which the observations are made up of trend, seasonal, cycle and regression components plus error. We go on to put Box–Jenkins ARIMA models into state space form, thus demonstrating that these models are special cases of state space models. Next we discuss the history of exponential smoothing and show how it relates to simple forms of state space and ARIMA models. We follow this by considering various aspects of regression with or without time-varying coefficients or autocorrelated errors. We also present a treatment of dynamic factor analysis. Further topics discussed are simultaneous modelling series from different sources, benchmarking, continuous time models and spline smoothing in discrete and continuous time. These considerations apply to minimum variance linear unbiased systems and to Bayesian treatments as well as to classical models.

Chapter 4 begins with a set of four lemmas from elementary multivariate regression which provides the essentials of the theory for the general linear state space model from both a classical and a Bayesian standpoint. These have the useful property that they produce the same results for Gaussian assumptions of the model and for linear minimum variance criteria where the Gaussian assumptions are dropped. The implication of these results is that we only need to prove formulae for classical models assuming normality and they remain valid for linear minimum variance and for Bayesian assumptions. The four lemmas lead to derivations of the Kalman filter and smoothing recursions for the estimation of the state vector and its conditional variance matrix given the data. We also derive recursions for estimating the observation and state disturbances. We derive the simulation smoother which is an important tool in the simulation methods we employ later in the book. We show that allowance for missing observations and forecasting are easily dealt with in the state space framework.

Computational algorithms in state space analyses are mainly based on recursions, that is, formulae in which we calculate the value at time $t+1$ from earlier values for $t, t-1, \dots, 1$. The question of how these recursions are started up at the beginning of the series is called initialisation; it is dealt with in Chapter 5. We give a general treatment in which some elements of the initial state vector have known distributions while others are diffuse, that is, treated as random variables with infinite variance, or are treated as unknown constants to be estimated by maximum likelihood.

Chapter 6 discusses further computational aspects of filtering and smoothing and begins by considering the estimation of a regression component of the model and intervention components. It next considers the square root filter and smoother which may be used when the Kalman filter and smoother show signs of numerical instability. It goes on to discuss how multivariate time series can be treated as univariate series by bringing elements of the observational vectors into the system one at a time, with computational savings relative to the multivariate treatment in some cases. Further modifications are discussed where the observation vector is high-dimensional. The chapter concludes by discussing computer algorithms.

In Chapter 7, maximum likelihood estimation of parameters is considered both for the case where the distribution of the initial state vector is known and for the case where at least some elements of the vector are diffuse or are treated as fixed and unknown. The use of the score vector and the EM algorithm is discussed. The effect of parameter estimation on variance estimation is examined.

Up to this point the exposition has been based on the classical approach to inference in which formulae are worked out on the assumption that parameters are known, while in applications unknown parameter values are replaced by appropriate estimates. In Bayesian analysis the parameters are treated as random variables with a specified or a noninformative prior joint density which necessitates treatment by simulation techniques which are not introduced until Chapter 13. Chapter 13 partly considers a Bayesian analysis of the linear Gaussian model both for the case where the prior density is proper and for the case where it is noninformative. We give formulae from which the posterior mean can be calculated for functions of the state vector, either by numerical integration or by simulation. We restrict attention to functions which, for given values of the parameters, can be calculated by the Kalman filter and smoother.

In Chapter 8 we illustrate the use of the methodology by applying the techniques that have been developed to a number of analyses based on real data. These include a study of the effect of the seat belt law on road accidents in Great Britain, forecasting the number of users logged on to an Internet server, fitting acceleration against time for a simulated motorcycle accident and a dynamic factor analysis for the term structure of US interest rates.

1.3 Non-Gaussian and nonlinear models

Part II of the book extends the treatment to state space models which are not both linear and Gaussian. Chapter 9 illustrates the range of non-Gaussian and nonlinear models that can be analysed using the methods of Part II. This includes exponential family models such as the Poisson distribution for the conditional distribution of the observations given the state. It also includes heavy-tailed distributions for the observational and state disturbances, such as the t-distribution and mixtures of normal densities. Departures from linearity of the models are studied for cases where the basic state space structure is preserved. Financial models such as stochastic volatility models are investigated from the state space point of view.

Chapter 10 considers approximate methods for analysis of non-Gaussian and nonlinear models, that is, extended Kalman filter methods and unscented methods. It also discusses approximate methods based on first and second order Taylor expansions. We show how to calculate the conditional mode of the state given the observations for the non-Gaussian model by iterated use of the Kalman filter and smoother. We then find the linear Gaussian model with the same conditional mode given the observations.

The simulation techniques for exact handling of non-Gaussian and nonlinear models are based on importance sampling and are described in Chapter 11. We then find the linear Gaussian model with the same conditional mode given the observations. We use the conditional density of the state given the observations for an approximating linear Gaussian model as the importance density. We draw random samples from this density for the simulation using the simulation smoother described in Chapter 4. To improve efficiency we introduce two antithetic variables intended to balance the simulation sample for location and scale.

In Chapter 12 we emphasise the fact that simulation for time series can be done sequentially, that is, instead of selecting an entire new sample for each time point t, which is the method suggested in Section 12.2, we fix the sample at the values previously obtained at time $\ldots, t-2, t-1$, and choose a new value at time t only. New recursions are required for the resulting simulations. This method is called particle filtering.

In Chapter 13 we discuss the use of importance sampling for the estimation of parameters in Bayesian analysis for models of Part I and Part II. An alternative simulation technique is Markov chain Monte Carlo. We prefer to use importance sampling for the problems considered in this book but a brief description is given for comparative purposes.

We provide examples in Chapter 14 which illustrate the methods that have been developed in Part II for analysing observations using non-Gaussian and nonlinear state space models. The illustrations include the monthly number of van drivers killed in road accidents in Great Britain, outlying observations in quarterly gas consumption, the volatility of exchange rate returns and analysis of the results of the annual boat race between teams of the universities of Oxford and Cambridge.

1.4 Prior knowledge

Only basic knowledge of statistics and matrix algebra is needed in order to understand the theory in this book. In statistics, an elementary knowledge is required of the conditional distribution of a vector y given a vector x in a multivariate normal distribution; the central results needed from this area for much of the theory of the book are stated in the lemmas in Section 4.2. Little previous knowledge of time series analysis is required beyond an understanding of the concepts of a stationary time series and the autocorrelation function. In matrix algebra all that is needed are matrix multiplication and inversion of matrices, together with basic concepts such as rank and trace.

1.5 Notation

Although a large number of mathematical symbols are required for the exposition of the theory in this book, we decided to confine ourselves to the standard

English and Greek alphabets. The effect of this is that we occasionally need to use the same symbol more than once; we have aimed however at ensuring that the meaning of the symbol is always clear from the context. We present below a list of the main conventions we have employed.

- The same symbol 0 is used to denote zero, a vector of zeros or a matrix of zeros.
- The symbol I_k denotes an identity matrix of dimension k.
- We use the generic notation $p(\cdot)$, $p(\cdot,\cdot)$, $p(\cdot|\cdot)$ to denote a probability density, a joint probability density and a conditional probability density.
- If x is a random vector with μ and variance matrix V and which is not necessarily normal, we write $x \sim (\mu, V)$.
- If x is a random vector which is normally distributed with mean vector μ and variance matrix V, we write $x \sim \mathrm{N}(\mu, V)$.
- If x is a random variable with the chi-squared distribution with ν degrees of freedom, we write $x \sim \chi^2_\nu$.
- We use the same symbol $\mathrm{Var}(x)$ to denote the variance of a scalar random variable x and the variance matrix of a random vector x.
- We use the same symbol $\mathrm{Cov}(x, y)$ to denote the covariance between scalar random variables x and y, between a scalar random variable x and a random vector y, and between random vectors x and y.
- The symbol $\mathrm{E}(x|y)$ denotes the conditional expectation of x given y; similarly for $\mathrm{Var}(x|y)$ and $\mathrm{Cov}(x, y|z)$ for random vectors x, y and z.
- The symbol $\mathrm{diag}(a_1, \ldots, a_k)$ denotes the $\ell \times \ell$ matrix with nonsingular matrix elements $a_1, \ldots, a_k$ down the leading diagonal and zeros elsewhere where $\ell = \sum_{i=1}^k \mathrm{rank}(a_i)$.

1.6 Other books on state space methods

Without claiming complete coverage, we list here a number of books which contain treatments of state space methods.

First we mention three early books written from an engineering standpoint: Jazwinski (1970), Sage and Melsa (1971) and Anderson and Moore (1979). A later book from a related standpoint is Young (1984).

Books written from the standpoint of statistics and econometrics include Harvey (1989), who gives a comprehensive state space treatment of structural time series models together with related state space material, West and Harrison (1997), who give a Bayesian treatment with emphasis on forecasting, Kitagawa and Gersch (1996) and Kim and Nelson (1999). A complete Bayesian treatment for specific classes of time series models including the state space model is given by Frühwirth-Schnatter (2006). Fundamental and rigorous statistical treatments of classes of hidden Markov models, which include our nonlinear non-Gaussian state space model of Part II, is presented in Cappé, Moulines and Rydén

(2005). An introductory and elementary treatment of state space methods from a practioners' perspective is provided by Commandeur and Koopman (2007).

More general books on time series analysis and related topics which cover partial treatments of state space topics include Brockwell and Davis (1987) (39 pages on state space out of about 570), Chatfield (2003) (14 pages out of about 300), Harvey (1993) (48 pages out of about 300), Hamilton (1994) (37 pages on state space out of about 800 pages) and Shumway and Stoffer (2000) (112 pages out of about 545 pages). The monograph of Jones (1993) on longitudinal models has three chapters on state space (66 pages out of about 225). The book by Fahrmeir and Tutz (1994) on multivariate analysis based on generalised linear modelling has a chapter on state space models (48 pages out of about 420). Finally, the book by Teräsvirta, Tjostheim and Granger (2011) on nonlinear time series modelling has one chapter on the treatment of nonlinear state space models (32 pages out of about 500 pages).

Books on time series analysis and similar topics with minor treatments of state space analysis include Granger and Newbold (1986) and Mills (1993). We mention finally the book edited by Doucet, De Freitas and Gordon (2001) which contains a collection of articles on Monte Carlo (particle) filtering and the book edited by Akaike and Kitagawa (1999) which contains 6 chapters (88 pages) on illustrations of state space analysis out of a total of 22 chapters (385 pages).

1.7 Website for the book

We will maintain a website for the book at

`http://www.ssfpack.com/dkbook.html`

for data, code, corrections and other relevant information. We will be grateful to readers if they inform us about their comments and errors in the book so corrections can be placed on the site.