

Use R!

Series Editors:

Robert Gentleman Kurt Hornik Giovanni Parmigiani

For other titles published in this series, go to
<http://www.springer.com/series/6991>

Giovanni Petris • Sonia Petrone • Patrizia Campagnoli

Dynamic Linear Models with R

 Springer

Giovanni Petris
Department of Mathematical Sciences
University of Arkansas
Fayetteville, AR 72701
USA
gpetris@uark.edu

Sonia Petrone
Department of Decision Sciences
Bocconi University
Via Roentgen, 1
20136 Milano
Italy
sonia.petrone@unibocconi.it

Patrizia Campagnoli
Department of Decision Sciences
Bocconi University
Via Roentgen, 1
20136 Milano
Italy
patrizia.campagnoli@unibocconi.it

ISBN 978-0-387-77237-0 e-ISBN 978-0-387-77238-7
DOI 10.1007/b135794
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2009926480

© Springer Science+Business Media, LLC 2009

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

A Mary e Giulio
G.P.

A Francesca, ai miei nipoti
S.P.

A Andrea e Michele
P.C.

Preface

This book aims at introducing the reader to statistical time series analysis by dynamic linear models. We have tried to be precise and rigorous in discussing the main concepts and tools, yet keeping a simple and friendly style of presentation. The main methods and models are widely illustrated with examples based on real data, implemented in R. Together with the book, we developed an R package for inference and forecasting with dynamic linear models; the `dlm` package is available as a contributed package in the Comprehensive R Archive Network at <http://www.r-project.org/>.

In the recent years, there has been an enormous growth of interest for statistical applications of dynamic linear models and, more generally, state-space models, in a wide range of applied fields, such as biology, economics, finance, marketing, quality control, engineering, demography, climatology, to mention only a few. State space models provide a very flexible yet fairly simple tool for analyzing dynamic phenomena and evolving systems, and have significantly contributed to extend the classical domains of application of statistical time series analysis to non-stationary, irregular processes, to systems evolving in continuous-time, to multivariate, continuous and discrete data. An extremely wide range of applied problems can be treated inside the framework of dynamic linear models or, more generally, state-space models.

The book covers the basic notions of dynamic linear models and state space models, the celebrated Kalman filter for estimation and forecasting in a dynamic linear model with known parameters, and maximum likelihood estimation. It also presents a wide array of specific dynamic linear models particularly suited for time series analysis, both for univariate and multivariate data. But these topics are of course also covered in other very good books in the rich literature on dynamic linear models, and several statistical softwares include packages for time series analysis through maximum likelihood and Kalman filtering. What we felt was somehow missing was an up to date, rigorous yet friendly reference—and software—for applied Bayesian time series analysis through dynamic linear models and state space models. This seemed to be missing despite the fact that the Bayesian approach has become

more and more popular in applications, due to the availability of modern and efficient computational tools. So, while also covering maximum likelihood methods, our focus in the book is on *Bayesian* time series analysis based on dynamic linear models.

We do not expect the reader to be an expert in Bayesian inference, so we begin with a short introduction to the Bayesian approach in Chapter 1. Also for a Bayesian reader, this is useful to set the notation and to underline some basic concepts that are used in the following chapters: for example, in presenting the simplest notions, such as Bayesian conjugate inference for a Gaussian model, we underline the recursive structure of the estimates, that will be one of the basic aspects of inference for dynamic linear models. Chapter 2 introduces the general setting of state space models and dynamic linear models, including the fundamental algorithms to sequentially update estimates and forecasts and the Kalman filter. Chapters 3 and 4 are in a sense the core of the book. In Chapter 3 the reader will find a discussion of a broad spectrum of specific models suited for the analysis of many kinds of data showing different features. Thus, Chapter 3 should be considered as a toolbox, illustrating a set of models from which the user can select the most appropriate for the application at hand. Chapter 4 covers maximum likelihood and Bayesian inference for dynamic linear models containing unknown parameters—which is always the case in practice. Many of the models introduced in Chapter 3 are discussed again there in this perspective. For most of the covered models we provide detailed examples of their use, correlated with the relevant R code. When possible, Bayesian estimates are evaluated using closed form algorithms. But in more elaborate models, analytical computations become intractable and simulation techniques are used to approximate the Bayesian solutions. We describe Markov chain Monte Carlo methods for Bayesian inference in dynamic linear models. The R package `dlm` provides functions for one of the basic steps in Bayesian computations in dynamic linear models, the so-called *forward filtering-backward sampling* algorithm, and other computational tools, with many examples, are provided. In Chapter 5, we present modern sequential Monte Carlo and particle filter algorithms for on-line estimation and forecasting.

Of course we cannot cover all of the extremely rich variety of models, applications, and problems in Bayesian inference with dynamic linear models, and many things will be missing. However, we hope to give a solid background on the main concepts and notions, leading the reader to acquire the skills for specific, personal elaborations, for which the flexibility of R and the `dlm` package will provide convenient, helpful tools. On the web site of the book, definetti.uark.edu/~gpetris/dlm, the reader will find data sets not included in the package and the code to run all the examples in the book. In addition, we plan to post there an updated list of errata.

The motivation for this book came from the authors' teaching experience in courses on time series analysis. We wanted to teach a course including—besides the classical ARMA models, descriptive techniques, exponential smoothing, and so on—more modern approaches, in particular Bayesian inference for time series through dynamic linear models. Again, we felt that a textbook, and a friendly but flexible software, were missing. So we started working on this project. We hope students, researchers, and practitioners will find the book and the software that resulted from our effort of some help.

We would like to thank Springer-Verlag's referees for their encouragement and valuable suggestions. Our thanks go also to our editor, John Kimmel, for his patience and support.

The `dlm` package would not exist without R, for which we thank R-core. Several people on `r-help`, the general R mailing list, have contributed their suggestions and feedback during the development of the package: we thank all of them. In particular, we thank Spencer Graves and Michael Lavine for their comments and suggestions on earlier versions of the package. Michael Lavine taught a course at the University of Massachusetts using R and `dlm` from an early draft of the book, and we thank him for the valuable feedback he gave us. One of the authors (GP) taught some short courses based on preliminary versions of the book at Bocconi University and the University of Roma 3 and would like to thank Pietro Muliere, Carlo Favero, Julia Mortera, and his coauthor, Sonia Petrone, for the kind invitations and the hospitality. SP used draft versions of the book in her graduate courses on time series analysis at Bocconi University: students' feedback has been precious. We thank all our students at the University of Arkansas, Bocconi University, and the University of Roma 3 who, with their comments, questions, suggestions, interest and enthusiasm, have contributed to the development of this book. Among them, a special thanks goes to Paolo Bonomolo and Guido Morandini.

Needless to say, the responsibility for any remaining mistakes, obscurities, or omissions—in the book and in the package—lies solely with us.

Fayetteville, Arkansas
and
Milano, Italy
December 15, 2008

Giovanni Petris
Sonia Petrone
Patrizia Campagnoli

Contents

1	Introduction: basic notions about Bayesian inference	1
1.1	Basic notions	2
1.2	Simple dependence structures	5
1.3	Synthesis of conditional distributions	11
1.4	Choice of the prior distribution	14
1.5	Bayesian inference in the linear regression model	18
1.6	Markov chain Monte Carlo methods	22
1.6.1	Gibbs sampler	24
1.6.2	Metropolis–Hastings algorithm	24
1.6.3	Adaptive rejection Metropolis sampling	25
	Problems	29
2	Dynamic linear models	31
2.1	Introduction	31
2.2	A simple example	35
2.3	State space models	39
2.4	Dynamic linear models	41
2.5	Dynamic linear models in package <code>dlm</code>	43
2.6	Examples of nonlinear and non-Gaussian state space models	48
2.7	State estimation and forecasting	49
2.7.1	Filtering	51
2.7.2	Kalman filter for dynamic linear models	53
2.7.3	Filtering with missing observations	59
2.7.4	Smoothing	60
2.8	Forecasting	66
2.9	The innovation process and model checking	73
2.10	Controllability and observability of time-invariant DLMs	77
2.11	Filter stability	80
	Problems	83

3	Model specification	85
3.1	Classical tools for time series analysis	85
3.1.1	Empirical methods	85
3.1.2	ARIMA models	87
3.2	Univariate DLMs for time series analysis	88
3.2.1	Trend models	89
3.2.2	Seasonal factor models	100
3.2.3	Fourier form seasonal models	102
3.2.4	General periodic components	109
3.2.5	DLM representation of ARIMA models	112
3.2.6	Example: estimating the output gap	115
3.2.7	Regression models	121
3.3	Models for multivariate time series	125
3.3.1	DLMs for longitudinal data	126
3.3.2	Seemingly unrelated time series equations	127
3.3.3	Seemingly unrelated regression models	132
3.3.4	Hierarchical DLMs	134
3.3.5	Dynamic regression	136
3.3.6	Common factors	138
3.3.7	Multivariate ARMA models	139
	Problems	142
4	Models with unknown parameters	143
4.1	Maximum likelihood estimation	144
4.2	Bayesian inference	148
4.3	Conjugate Bayesian inference	149
4.3.1	Unknown covariance matrices: conjugate inference	150
4.3.2	Specification of W_t by discount factors	152
4.3.3	A discount factor model for time-varying V_t	158
4.4	Simulation-based Bayesian inference	160
4.4.1	Drawing the states given $y_{1:T}$: forward filtering backward sampling	161
4.4.2	General strategies for MCMC	162
4.4.3	Illustration: Gibbs sampling for a local level model	165
4.5	Unknown variances	167
4.5.1	Constant unknown variances: d Inverse Gamma prior	167
4.5.2	Multivariate extensions	171
4.5.3	A model for outliers and structural breaks	177
4.6	Further examples	186
4.6.1	Estimating the output gap: Bayesian inference	186
4.6.2	Dynamic regression	192
4.6.3	Factor models	200
	Problems	206

5	Sequential Monte Carlo methods	207
5.1	The basic particle filter	208
5.1.1	A simple example	213
5.2	Auxiliary particle filter	216
5.3	Sequential Monte Carlo with unknown parameters	219
5.3.1	A simple example with unknown parameters	226
5.4	Concluding remarks	228
A	Useful distributions	231
B	Matrix algebra: Singular Value Decomposition	237
	Index	241
	References	245