

Time Series Analysis by State Space Methods

J. DURBIN

*Department of Statistics
London School of Economics and Political Science*

and

S. J. KOOPMAN

*Department of Econometrics
Free University Amsterdam*

Wendt

QA

280

DP7

2001



OXFORD
UNIVERSITY PRESS

Contents

1	Introduction	1
1.1	Basic ideas of state space analysis	1
1.2	Linear Gaussian model	1
1.3	Non-Gaussian and nonlinear models	3
1.4	Prior knowledge	4
1.5	Notation	4
1.6	Other books on state space methods	5
1.7	Website for the book	5

I THE LINEAR GAUSSIAN STATE SPACE MODEL

2	Local level model	9
2.1	Introduction	9
2.2	Filtering	11
2.2.1	The Kalman Filter	11
2.2.2	Illustration	12
2.3	Forecast errors	13
2.3.1	Cholesky decomposition	14
2.3.2	Error recursions	15
2.4	State smoothing	16
2.4.1	Smoothed state	16
2.4.2	Smoothed state variance	17
2.4.3	Illustration	18
2.5	Disturbance smoothing	19
2.5.1	Smoothed observation disturbances	20
2.5.2	Smoothed state disturbances	20
2.5.3	Illustration	21
2.5.4	Cholesky decomposition and smoothing	22
2.6	Simulation	22
2.6.1	Illustration	23
2.7	Missing observations	23
2.7.1	Illustration	25
2.8	Forecasting	25
2.8.1	Illustration	27

2.9	Initialisation	27
2.10	Parameter estimation	30
2.10.1	Loglikelihood evaluation	30
2.10.2	Concentration of loglikelihood	31
2.10.3	Illustration	32
2.11	Steady state	32
2.12	Diagnostic checking	33
2.12.1	Diagnostic tests for forecast errors	33
2.12.2	Detection of outliers and structural breaks	35
2.12.3	Illustration	35
2.13	Appendix: Lemma in multivariate normal regression	36
3	Linear Gaussian state space models	38
3.1	Introduction	38
3.2	Structural time series models	39
3.2.1	Univariate models	39
3.2.2	Multivariate models	44
3.2.3	STAMP	45
3.3	ARMA models and ARIMA models	46
3.4	Exponential smoothing	49
3.5	State space versus Box-Jenkins approaches	51
3.6	Regression with time-varying coefficients	54
3.7	Regression with ARMA errors	54
3.8	Benchmarking	54
3.9	Simultaneous modelling of series from different sources	56
3.10	State space models in continuous time	57
3.10.1	Local level model	57
3.10.2	Local linear trend model	59
3.11	Spline smoothing	61
3.11.1	Spline smoothing in discrete time	61
3.11.2	Spline smoothing in continuous time	62
4	Filtering, smoothing and forecasting	64
4.1	Introduction	64
4.2	Filtering	65
4.2.1	Derivation of Kalman filter	65
4.2.2	Kalman filter recursion	67
4.2.3	Steady state	68
4.2.4	State estimation errors and forecast errors	68
4.3	State smoothing	70
4.3.1	Smoothed state vector	70
4.3.2	Smoothed state variance matrix	72
4.3.3	State smoothing recursion	73
4.4	Disturbance smoothing	73

4.4.1	Smoothed disturbances	73
4.4.2	Fast state smoothing	75
4.4.3	Smoothed disturbance variance matrices	75
4.4.4	Disturbance smoothing recursion	76
4.5	Covariance matrices of smoothed estimators	77
4.6	Weight functions	81
4.6.1	Introduction	81
4.6.2	Filtering weights	81
4.6.3	Smoothing weights	82
4.7	Simulation smoothing	83
4.7.1	Simulating observation disturbances	84
4.7.2	Derivation of simulation smoother for observation disturbances	87
4.7.3	Simulation smoothing recursion	89
4.7.4	Simulating state disturbances	90
4.7.5	Simulating state vectors	91
4.7.6	Simulating multiple samples	92
4.8	Missing observations	92
4.9	Forecasting	93
4.10	Dimensionality of observational vector	94
4.11	General matrix form for filtering and smoothing	95
5	Initialisation of filter and smoother	99
5.1	Introduction	99
5.2	The exact initial Kalman filter	101
5.2.1	The basic recursions	101
5.2.2	Transition to the usual Kalman filter	104
5.2.3	A convenient representation	105
5.3	Exact initial state smoothing	106
5.3.1	Smoothed mean of state vector	106
5.3.2	Smoothed variance of state vector	107
5.4	Exact initial disturbance smoothing	109
5.5	Exact initial simulation smoothing	110
5.6	Examples of initial conditions for some models	110
5.6.1	Structural time series models	110
5.6.2	Stationary ARMA models	111
5.6.3	Nonstationary ARIMA models	112
5.6.4	Regression model with ARMA errors	114
5.6.5	Spline smoothing	115
5.7	Augmented Kalman filter and smoother	115
5.7.1	Introduction	115
5.7.2	Augmented Kalman filter	115
5.7.3	Filtering based on the augmented Kalman filter	116

5.7.4	Illustration: the local linear trend model	118
5.7.5	Comparisons of computational efficiency	119
5.7.6	Smoothing based on the augmented Kalman filter	120
6	Further computational aspects	121
6.1	Introduction	121
6.2	Regression estimation	121
6.2.1	Introduction	121
6.2.2	Inclusion of coefficient vector in state vector	122
6.2.3	Regression estimation by augmentation	122
6.2.4	Least squares and recursive residuals	123
6.3	Square root filter and smoother	124
6.3.1	Introduction	124
6.3.2	Square root form of variance updating	125
6.3.3	Givens rotations	126
6.3.4	Square root smoothing	127
6.3.5	Square root filtering and initialisation	127
6.3.6	Illustration: local linear trend model	128
6.4	Univariate treatment of multivariate series	128
6.4.1	Introduction	128
6.4.2	Details of univariate treatment	129
6.4.3	Correlation between observation equations	131
6.4.4	Computational efficiency	132
6.4.5	Illustration: vector splines	133
6.5	Filtering and smoothing under linear restrictions	134
6.6	The algorithms of SsfPack	134
6.6.1	Introduction	134
6.6.2	The SsfPack function	135
6.6.3	Illustration: spline smoothing	136
7	Maximum likelihood estimation	138
7.1	Introduction	138
7.2	Likelihood evaluation	138
7.2.1	Loglikelihood when initial conditions are known	138
7.2.2	Diffuse loglikelihood	139
7.2.3	Diffuse loglikelihood evaluated via augmented Kalman filter	140
7.2.4	Likelihood when elements of initial state vector are fixed but unknown	141
7.3	Parameter estimation	142
7.3.1	Introduction	142
7.3.2	Numerical maximisation algorithms	142
7.3.3	The score vector	144
7.3.4	The EM algorithm	147

7.3.5	Parameter estimation when dealing with diffuse initial conditions	149
7.3.6	Large sample distribution of maximum likelihood estimates	150
7.3.7	Effect of errors in parameter estimation	150
7.4	Goodness of fit	152
7.5	Diagnostic checking	152
8	Bayesian analysis	155
8.1	Introduction	155
8.2	Posterior analysis of state vector	155
8.2.1	Posterior analysis conditional on parameter vector	155
8.2.2	Posterior analysis when parameter vector is unknown	155
8.2.3	Non-informative priors	158
8.3	Markov chain Monte Carlo methods	159
9	Illustrations of the use of the linear Gaussian model	161
9.1	Introduction	161
9.2	Structural time series models	161
9.3	Bivariate structural time series analysis	167
9.4	Box-Jenkins analysis	169
9.5	Spline smoothing	172
9.6	Approximate methods for modelling volatility	175

II NON-GAUSSIAN AND NONLINEAR STATE SPACE MODELS

10	Non-Gaussian and nonlinear state space models	179
10.1	Introduction	179
10.2	The general non-Gaussian model	179
10.3	Exponential family models	180
10.3.1	Poisson density	181
10.3.2	Binary density	181
10.3.3	Binomial density	181
10.3.4	Negative binomial density	182
10.3.5	Multinomial density	182
10.4	Heavy-tailed distributions	183
10.4.1	t -Distribution	183
10.4.2	Mixture of normals	184
10.4.3	General error distribution	184
10.5	Nonlinear models	184
10.6	Financial models	185
10.6.1	Stochastic volatility models	185

10.6.2	General autoregressive conditional heteroscedasticity	187
10.6.3	Durations: exponential distribution	188
10.6.4	Trade frequencies: Poisson distribution	188
11	Importance sampling	189
11.1	Introduction	189
11.2	Basic ideas of importance sampling	190
11.3	Linear Gaussian approximating models	191
11.4	Linearisation based on first two derivatives	193
11.4.1	Exponential family models	195
11.4.2	Stochastic volatility model	195
11.5	Linearisation based on the first derivative	195
11.5.1	t -distribution	197
11.5.2	Mixture of normals	197
11.5.3	General error distribution	197
11.6	Linearisation for non-Gaussian state components	198
11.6.1	t -distribution for state errors	199
11.7	Linearisation for nonlinear models	199
11.7.1	Multiplicative models	201
11.8	Estimating the conditional mode	202
11.9	Computational aspects of importance sampling	204
11.9.1	Introduction	204
11.9.2	Practical implementation of importance sampling	204
11.9.3	Antithetic variables	205
11.9.4	Diffuse initialisation	206
11.9.5	Treatment of t -distribution without importance sampling	208
11.9.6	Treatment of Gaussian mixture distributions without importance sampling	210
12	Analysis from a classical standpoint	212
12.1	Introduction	212
12.2	Estimating conditional means and variances	212
12.3	Estimating conditional densities and distribution functions	213
12.4	Forecasting and estimating with missing observations	214
12.5	Parameter estimation	215
12.5.1	Introduction	215
12.5.2	Estimation of likelihood	215
12.5.3	Maximisation of loglikelihood	216
12.5.4	Variance matrix of maximum likelihood estimate	217
12.5.5	Effect of errors in parameter estimation	217

12.5.6	Mean square error matrix due to simulation	217
12.5.7	Estimation when the state disturbances are Gaussian	219
12.5.8	Control variables	219
13	Analysis from a Bayesian standpoint	222
13.1	Introduction	222
13.2	Posterior analysis of functions of the state vector	222
13.3	Computational aspects of Bayesian analysis	225
13.4	Posterior analysis of parameter vector	226
13.5	Markov chain Monte Carlo methods	228
14	Non-Gaussian and nonlinear illustrations	230
14.1	Introduction	230
14.2	Poisson density: van drivers killed in Great Britain	230
14.3	Heavy-tailed density: outlier in gas consumption in UK	233
14.4	Volatility: pound/dollar daily exchange rates	236
14.5	Binary density: Oxford–Cambridge boat race	237
14.6	Non-Gaussian and nonlinear analysis using SsfPack	238
	References	241
	Author index	249
	Subject index	251

ILLUSTRATION SAMPLE

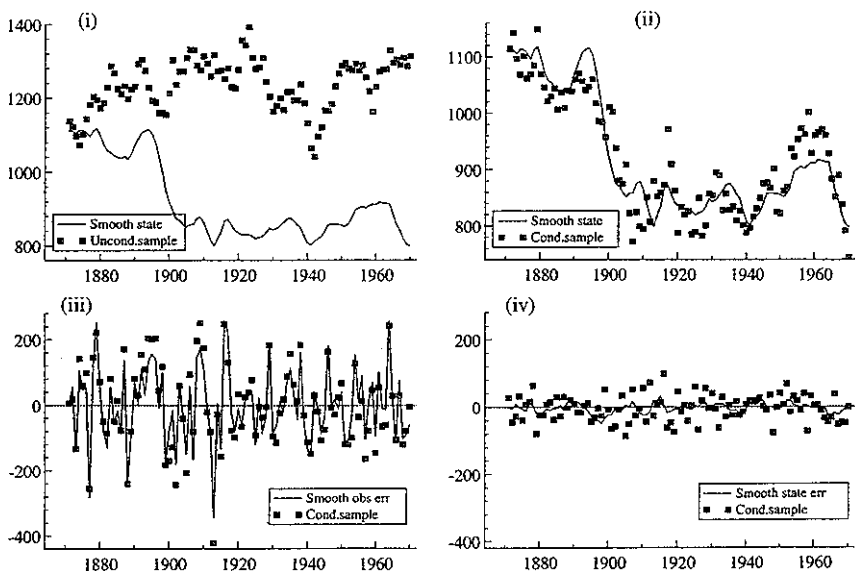


Fig. 2.4. Simulation: (i) smoothed state $\hat{\alpha}_t$ and sample $\alpha_t^{(i)}$; (ii) smoothed state $\hat{\alpha}_t$ and sample $\tilde{\alpha}_t$; (iii) smoothed observation error $\hat{\varepsilon}_t$ and sample $\tilde{\varepsilon}_t$; (iv) smoothed state error $\hat{\eta}_t$ and sample $\tilde{\eta}_t$.

It is, however, easier and more transparent to proceed as follows, using the original time domain. For filtering at times $t = \tau, \dots, \tau^* - 1$, we have

$$E(\alpha_t | Y_{t-1}) = E(\alpha_t | Y_{\tau-1}) = E\left(\alpha_\tau + \sum_{j=\tau}^{t-1} \eta_j \middle| Y_{\tau-1}\right) = a_\tau$$

and

$$\text{Var}(\alpha_t | Y_{t-1}) = \text{Var}(\alpha_t | Y_{\tau-1}) = \text{Var}\left(\alpha_\tau + \sum_{j=\tau}^{t-1} \eta_j \middle| Y_{\tau-1}\right) = P_\tau + (t - \tau)\sigma_\eta^2.$$

giving

$$a_{t+1} = a_t, \quad P_{t+1} = P_t + \sigma_\eta^2, \quad t = \tau, \dots, \tau^* - 1, \quad (2.38)$$

the remaining values a_t and P_t being given as before by (2.11) for $t = 1, \dots, \tau$ and $t = \tau^*, \dots, n$. The consequence is that we can use the original filter (2.11) for all t by taking $v_t = 0$ and $K_t = 0$ at the missing time points. The same procedure is used when more than one group of observations is missing. It follows that allowing for missing observations when using the Kalman filter is extremely simple.

The forecast error recursions from which we derive the smoothing recursions are given by (2.18). These error-updating equations at the missing time points