State space models: a brief history and some recent developments

Ludwig Fahrmeir
University of Munich, Department of Statistics
Ludwigstr.33
D-80539 München, Germany
fahrmeir@stat.uni-muenchen.de

1. Introduction

State space models, also termed dynamic models, relate observations y_t, $t = 1, 2, ...$, on a response variable Y to unobserved "states" or "parameters" α_t, t=1,2,..., by an observation model for y_t given α_t. The states are assumed to follow a Markovian transition model. Gaussian linear state space models are defined by a linear observation model and a linear Markovian transition equation

(1) $$y_t = z_t' \alpha_t + \epsilon_t, \qquad t = 1, 2, ...$$

(2) $$\alpha_t = F_t \alpha_{t-1} + \xi_t, \quad t = 1, 2, ...$$

with independent i.i.d. sequences $\epsilon_t \sim N(0, \sigma^2)$, $\xi_t \sim N(0, Q)$, and an independent initial value $\alpha_0 \sim N(a_0, Q_0)$.

State space models have their origin in systems theory, and famous early applications in the Apollo and Polaris aerospace programs, see, e.g., Hutchinson(1984). In this context, the observation equation (1) describes radar observations y_t, disturbed by noise, on the state vector α_t (position, velocity,...) of a spacecraft, ship or vehicle, and the transition equation (2) is a linearized and discretized version of motion in space. Given the observations $y = (y_1, ...y_t)$, on-line estimation or *filtering* of α_t for t=1,2,..., and *prediction* of α_s, s>t, are of primary interest. Related problems arise in on-line monitoring of patients or in ecological processes (Smith and West, 1983; Frühwirth-Schnatter, 1994a). The main applications in statistics are structural time series and dynamic regression models of the form $y_t = \mu_t + \gamma_t + x_t' \beta_t + \epsilon_t$, with a trend function μ_t, a seasonal component γ_t and possibly time-varying effects of covariates x_t. Gathering μ_t, γ_t and β_t into a state vector and defining appropriate transition models (2), these models can be written in state space form, see, e.g., Harvey(1989). Given observations $y = (y_1, ..., y_t, ..., y_T)$, estimation or *smoothing* of the whole sequence $\alpha = (\alpha_1, ..., \alpha_t, ..., \alpha_T)$ is of interest.

Under quadratic loss functions, posterior mean *filters* $\alpha_{t|t} = E(\alpha_t | y_1, ..., y_t)$ or *smoothers* $\alpha_{t|T} = E(\alpha_t | y = (y_1, ..., y_T))$ are optimal estimators. If z_t and F_t in (1),(2), as well as σ^2,Q and a_0, Q_0 are known, the famous Kalman filter and smoother (KFS) provides an analytical solution: It computes the filter estimates $\alpha_{t|t}$, t=1,2,...,T, in forward recursions and the smoother $\alpha_{t|T}$, t=T-1,...,1, in backward steps. Moreover, due to linearity and normality assumptions, marginal and joint posterior distributions are Gaussian.

2. Function estimation via Gaussian models

In the following, we sketch the lines of arguments that correspond to the historically first derivation of a special KFS by Thiele in 1880 and show the close relationship to nonparametric function estimation. Consider first the classical smoothing problem, where observations $y = (y_1, ..., y_T)$ are assumed to be the sum

(3) $$y_t = \mu_t + \epsilon_t, \quad t = 1, ..., T$$

of a smooth regression curve $\{\mu_t\}$, evaluated at equally spaced design points t, and i.i.d. errors $\epsilon_t \sim N(0, \sigma^2)$. In a state space approach for estimating $\{\mu_t\}$,(3) is supplemented by Gaussian random walk models of first (RW(1)) or second (RW(2)) order:

$$\mu_t = \mu_{t-1} + \xi_t \quad \text{or} \quad \mu_t = 2\mu_{t-1} - \mu_{t-2} + \xi_t, \quad \xi_t \sim N(0, q^2). \tag{4}$$

From a Bayesian point of view, these random walk models define *smoothness priors* on first and second differences, $\nabla^1 \mu_t = \mu_t - \mu_{t-1}$ and $\nabla^2 \mu_t = \mu_t - 2\mu_{t-1} + \mu_{t-2}$, respectively, that help to regularize the estimation problem by putting a penalty on deviations from horizontal or straight lines. Since (3) and (4) can be put in state space form, the KFS can be applied for given variances σ^2 and q^2 to compute posterior means $\mu_{t|T} = E(\mu_t|y)$ as optimal smoothers, together with posterior variances. Since the posterior is Gaussian, mean and mode coincide, so that $\mu_{t|T}$, t=1,...,T, can also be obtained by maximizing the posterior. Assuming diffuse initial priors for μ_1, μ_2 and taking logarithms, this is equivalent to the classical optimal smoothing problem already considered by Whittaker(1923): Choose $\hat{\mu} = (\mu_{1|T}, ..., \mu_{T|T})$ as the minimizer of the penalized least squares criteria

$$\sum_{t=1}^{T} (y_t - \mu_t)^2 + \frac{\sigma^2}{q^2} \sum_{t=k+1}^{T} (\nabla^k \mu_t)^2, \qquad k = 1, 2, \tag{5}$$

for a RW(1) resp. a RW(2) prior. From (5), the close correspondence to spline smoothing becomes clear: The penalty terms are discretized versions of corresponding roughness penalties for quadratic or cubic smoothing splines, and the variance ratio $\lambda = \sigma^2/q^2$ acts as a smoothness parameter. This equivalence remains also valid for general Gaussian state space models (1),(2), see e.g. Fahrmeir and Tutz (1994, ch.8.1). Adopting the Bayesian smoothness priors approach offers additional possibilities for data driven choice of the smoothing parameter $\lambda = \sigma^2/q^2$ by estimating the variances via the ML principle or, in a fully Bayesian approach, by Markov chain Monte Carlo (Frühwirth-Schnatter, 1994; Carter and Kohn, 1994). We also note that the whole approach can be extended for unequally spaced observations. With such modifications, the state space approach can be used for Bayesian nonparametric function estimation in a regression model $y_i = \mu(x_i) + \epsilon_i$, i=1,...,n, with Gaussian errors. Here x can be any metrical covariate instead of time t, and the *ordered* covariate observations $x_{(1)} < ... < x_{(i)} < ... < x_{(n)}$ correspond to unequally space time points. Extensions to additive models $y_i = \mu_1(x_{i_1}) + ... + \mu_p(x_{i_p}) + \epsilon_i$ with several covariates $x_1, ..., x_p$ are not straightforward, however, because state space models and the KFS can only deal with one time scale or covariate. Therefore, additive models either require "Bayesian backfitting" (Hastie and Tibshirani, 1998) via Gibbs sampling, or other MCMC approaches based on state space models as in Fahrmeir and Lang (1998).

3. Function estimation via non-Gaussian models

Non-Gaussian and nonlinear state space models are obtained by dropping normality and linearity assumptions in (1) or (2). Although general filtering and smoothing integral recursions can be obtained in principle, exact calculations as with the KFS are no longer possible. Apart from comparably crude approximations like the extended Kalman filter, several methods have been proposed more recently: numerical (Kitagawa, 1987, for low-dimensional α_t's) or Monte Carlo filters (Kitagawa, 1998; Huerzeler and Kuensch, 1998; Pitt and Shephard, 1998), posterior mode estimation (Fahrmeir and Wagenpfeil, 1997; Fahrmeir and Tutz, 1994, ch.8) and fully Bayesian smoothing using Markov chain Monte Carlo simulation (Knorr-Held, 1998; Shephard and Pitt, 1997).

An important subclass are *dynamic generalized linear models* (e.g. Fahrmeir and Tutz, 1994, ch.8). They are obtained by assuming that observations y_t given α_t come from an exponential family density with mean

$$\mu_t = E(y_t|\alpha_t) = h(z_t'\alpha_t),$$

where h is some link function. If we retain the Gaussian transition model(2), the posterior mode estimate $\hat{\alpha} = (\hat{\alpha_1}, ..., \hat{\alpha_T})$ can be shown to maximize the penalized likelihood criterion $\sum l_t(\alpha_t) + PEN(\alpha)$, where $l_t(\alpha_t)$ is the likelihood contribution of $y_t|\alpha_t$, and PEN(α) is a roughness penalty as for example in (5). The posterior mode smoother can still be obtained by an iterative KFS, but it is no longer equivalent to the posterior mean obtained from MCMC output or other Monte Carlo methods. The smoothness priors approach based on dynamic generalized linear models can be used for fully *Bayesian function estimation* in *generalized additive regression models* for non-Gaussian responses y_i, covariate observations $x_{i_1}, ..., x_{i_p}$, and predictor $\eta_i = f_1(x_{i_1}) + ... + f_p(x_{i_p})$, i=1,...,n, see Fahrmeir and Lang (1998). This approach is also tailored for incorporation of exchangeable or spatial random effects to account for unobserved individual or spatial heterogeneity.

A simple modification of (4) or (5) allows for estimation of *unsmooth functions* $\{\mu_t\}$, such as the stylized functions Blocks, Bumps, Heavy Sine and Doppler constructed for wavelet shrinkage by Donoho and Johnstone (1994). The basic idea is to replace the constant variance q^2=var(ξ_t) in (4) or (5) by locally varying variances q_t^2, thereby replacing the *global* smoothing parameter $\lambda = \sigma^2/q^2$ by a *local* smoothing parameter $\lambda_t = \sigma^2/q_t^2$. To estimate the variance function $\{q_t^2\}$ together with the unknown regression curve $\{\mu_t\}$, we reparametrize by $h_t = log(q_t^2)$ and add a second RW(1) or RW(2) Gaussian smoothness prior for $\{h_t\}$. Thus, the model consists of (3), (4) with q^2 replaced by $q_t^2 = exp(h_t)$ and smoothness priors for $\{h_t\}$. A fully Bayesian implementation with inverse Gamma priors for σ_η^2 and MCMC inference is given in Fronk and Fahrmeir (1998) for Gaussian observation models. Extensions for non-Gaussian observation models, such as generalized additive regression models with unsmooth regression functions will be considered in future research.

REFERENCES

Carter, C.K. and Kohn, R. (1994). On Gibbs sampling for state space models. Biometrika 81, 541-553.

Donoho, D.L. and Johnstone, I.M. (1994). Ideal spatial adaption by wavelet shrinkage. Biometrika 81, 425-455.

Fahrmeir, L. and Tutz, G. (1994, 1997). Multivariate Statistical Modelling based on Generalized Linear Models. Springer Series in Statistics, New York.

Fahrmeir, L. and Wagenpfeil, S. (1997). Penalized likelihood estimation and iterative Kalman smoothing for non-Gaussian dynamic regression models. Comp. Stat. Data Analysis 24, 295-320.

Fahrmeir, L. and Lang, S. (1998). Bayesian inference for generalized regression based on dynamic models. Universität München, Institut für Statistik, SFB 386 Discussion paper 134.

Fronk, E.-M. and Fahrmeir, L. (1998). Function estimation with locally adaptive dynamic models. Universität München, Institut für Statistik, SFB 386 Discussion paper 135.

Frühwirth-Schnatter, S. (1994a). Applied State Space Modelling of Non-Gaussian Time Series Using Integration-Based Kalman Filtering. Statistics and Computing 4, 259-269.

Frühwirth-Schnatter, S. (1994b). Data augmentation and dynamic linear models. Journal of Time Series Analysis 15, 183-202.

Harvey, A.C. (1989). Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge, Cambridge University Press.

Hastie, T. and Tibshirani, R. (1988). Bayesian Backfitting. Preprint, Dept. of Statistics, Stanford.

Hürzeler, M. and Künsch, H.R. (1998). Monte Carlo approximation for general state-space models. Journal of Computational and Graphical Statistics 7, 175-193.

Hutchinson, C.E. (1984). The Kalman Filter Applied to Aerospace and Electronic Systems: IEEE Trans.Aero.Elect.Syst.AES-20.500-504.

Kitagawa, G. (1987): Non-Gaussian State-space Modeling of Nonstationary Time Series (with comments); JASA 82, 1032-1063.

Kitagawa, G. (1998). A Self-Organizing State-Space Model. Journal of the American Statistical Association 93, 1203-1215.

Knorr-Held, L. (1998). Conditional Prior Proposals in Dynamic Models. To appear in Scandinavian Journal of Statistics.

Pitt, M.K. and Shephard, N. (1997). Filtering via simulation: auxiliary particle filters. Nuffied College, Oxford, preprint.

Shephard, N. and Pitt, M.K. (1997). Likelihood analysis of non-Gaussian measurement time series. Biometrika 84, 653-668.

Smith, A. and West, M. (1983). Monitoring Renal Transplants: an Application of the Multi-Process Kalman Filter. Biometrics 39, 867-878.

Whittaker, E.T. (1923). On a new method of graduation. Proceedings of Edingburgh Mathmatical Society 41, 63-75.

RÉSUMÉ

Les modèles à espaces d'états et le filtre de Kalman sont un cadre naturel pour estimer des effets dependant du temps. Ici, nous considérons ces modèles comme une approche semi-parametrique Bayésienne pour la regression generalisée, et nous montrons des relations avec la methode de vraisemblance penalisée.