Fronk, Fahrmeir:

Function estimation with locally adaptive dynamic models

Sonderforschungsbereich 386, Paper 135 (1998)

Online unter: http://epub.ub.uni-muenchen.de/

Function estimation with locally adaptive dynamic models

Eva–Maria Fronk and Ludwig Fahrmeir
Institut für Statistik, Universität München

Abstract

We present a nonparametric Bayesian method for fitting unsmooth functions which is based on a locally adaptive hierarchical extension of standard dynamic or state space models. The main idea is to introduce locally varying variances in the states equations and to add a further smoothness prior for this variance function. Estimation is fully Bayesian and carried out by recent MCMC techniques. The whole approach can be understood as an alternative to other nonparametric function estimators, such as local regression with local bandwidth or smoothing parameter selection. Performance is illustrated with simulated data, including unsmooth examples constructed for wavelet shrinkage, and by an application to CP6 sales data.

1 Introduction

Nonparametric methods for fitting smooth curves, such as kernel, local or spline regression are now widely available and accepted. However these methods can have bad performance when estimating unsmooth functions with jumps, edges or regions where have gained considerable interest, with local bandwidth selection in kernel or local regression (Fan and Gijbels, 1995)) or with wavelet shrinkage (Donoho and Johnstone, 1994) as prominent approaches.

In this paper, we present a Bayesian nonparametric method for estimation of unsmooth functions that is based on locally adaptive dynamic or state space models. Compared to standard Gaussian state space models (e.g. Harvey, 1989; Fahrmeir and Tutz, 1994), we allow for unknown locally varying variances of the errors in the state equation for the unknown function and add a further smoothness prior for the variance function (Section 2). These varying variances correspond to locally varying bandwidths or smoothing parameters in the non–Bayesian methods.

Estimation is fully Bayesian and uses recent Markov chain Monte Carlo techniques, combining Gibbs sampling and a Metropolis–Hastings algorithm of Knorr–Held (1998). Details are given in Section 3. Performance is illustrated in Section 4 with simulated data, including the unsmooth functions constructed for wavelet shrinkage by Donoho and Johnstone (1994), and by an application to sales data.
Although we focus here on a simple Gaussian observation model suitable for one–dimensional curve fitting, the basic idea can be adapted to more general settings. Some of the resulting extensions, for example to non–Gaussian observations or to surface estimation, are mentioned in the conclusions.

2 Locally adaptive dynamic models

Consider first the classical smoothing problem for a response variable, where observations $\mathbf{y} = (y_1, y_2, \ldots, y_T)$ are assumed to be the sum

$$y_t = \alpha_t + \epsilon_t, \qquad t = 1, \ldots, T \tag{1}$$

of a smooth trend function or regression curve, evaluated at the observation or design points t, and independent Gaussian errors $\epsilon_t \sim \mathrm{N}(0, \sigma_\epsilon^2)$. We denote the vector of the function evaluations by $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_T)$ and use the same symbol for the whole curve. For simplicity, we assume equidistant design points. Extensions to non–equally spaced designs are shortly outlined at the end of the section. In a standard dynamic or state modelling approach for estimation of the unknown function $\boldsymbol{\alpha}$, more exactly its evaluations α_t, the observation model (1) is supplemented by a Gaussian random walk of first order

$$\alpha_t \;=\; \alpha_{t-1} + u_t, \qquad\qquad u_t \sim \mathrm{N}(0; q^2) \tag{2}$$

or of second order

$$\alpha_t \;=\; 2\alpha_{t-1} - \alpha_{t-2} + u_t, \quad u_t \sim \mathrm{N}(0; q^2), \tag{3}$$

denoted as RW(1) respectively as RW(2). The errors u_t are mutually independent and independent of observation errors ϵ_t. In addition we will assume diffuse priors for initial values α_1, α_2. From a Bayesian point of view, the random walk models 2 and 3 define smoothness priors on first and second differences $\alpha_t - \alpha_{t-1} = u_t$ respectively $\alpha_t - 2\alpha_{t-1} + \alpha_{t-2} = u_t$ that help to regularise the estimation problem. For given variances σ_ϵ^2 and q^2 the famous linear Kalman filter and smoother computes the posterior means $\hat{\alpha}_t = \mathrm{E}(\alpha_t \mid y_1, \ldots, y_T)$ as optimal smoothers, together with posterior variances. Since the posterior is Gaussian, mean and mode coincide, and therefore the estimates $\hat{\alpha}_t$, $t = 1, \ldots, T$ can also be obtained by maximizing the posterior. Taking logarithms, this leads to the classical optimal smoothing problem already considered by Whittaker (1923): Choose $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_T)$ as the minimizer of

$$\sum_{t=1}^{T}(y_t - \alpha_t)^2 \;+\; \frac{\sigma_\epsilon^2}{q^2}\sum_{t=2}^{T}(\alpha_t - \alpha_{t-1})^2 \tag{4}$$

for model (2), and

$$\sum_{t=1}^{T}(y_t - \alpha_t)^2 \;+\; \frac{\sigma_\epsilon^2}{q^2}\sum_{t=3}^{T}(\alpha_t - 2\alpha_{t-1} + \alpha_{t-2})^2 \tag{5}$$

for model (3). From (4) and (5), the close correspondence to spline smoothing becomes clear: The ratio $\lambda = \sigma_\epsilon/q^2$ is a *global* smoothing parameter and the penalty

terms are the discretized versions of corresponding penalty terms for quadratic and cubic smoothing splines. Already with a moderate number of observations, estimates $\hat{\alpha}_t$ are practically undistinguishable from spline smoothing estimates.

The basic idea for estimation of non smooth functions is to replace the constant variance q^2 in (2) and (3) by locally varying variances q_t^2 that are considered as evaluations of a variance function $\mathbf{q}$. This corresponds to replacing the global smoothing parameter by a *local* smoothing parameter $\lambda_t = \sigma_\epsilon / q_t^2$. To estimate the unknown variance function automatically together with the unknown curve $\boldsymbol{\alpha}$, we reparametrize by

$$h_t = log(q_t^2) \Longleftrightarrow q_t^2 = \exp(h_t)$$

and add a second smoothness prior in form of first or second order differences for $\mathbf{h} = (h_1, \ldots, h_T)$, implying also a smoothness prior for $\mathbf{q} = (q_1^2, \ldots, q_T^2)$. Thus we will arrive at the following *locally adaptive dynamic models.*

Observation model for $\mathbf{y} = (y_1, \ldots, y_T)$:

$$y_t = \alpha_t + \epsilon_t, \quad \epsilon \sim \mathrm{N}(0, \sigma_\epsilon^2) \tag{6}$$

Smoothness priors for $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_T)$:

$$\alpha_t = \alpha_{t-1} + u_t, \quad \text{or} \quad \alpha_t = 2\alpha_{t-1} - \alpha_{t-2} + u_t \tag{7}$$

with $u_t \sim \mathrm{N}(0, \exp(h_t))$.

Smoothness priors for $\mathbf{h} = (h_1, h_2, \ldots, h_T)$:

$$h_t = h_{t-1} + u_t, \quad \text{or} \quad h_t = 2h_{t-1} - h_{t-2} + \eta_t, \quad \eta \sim \mathrm{N}(0, \sigma_\eta^2). \tag{8}$$

We assume mutually independent errors ϵ_t, u_t and η_t and diffuse priors for initial values α_1, α_2 h_1, h_2. The model definition is completed by the common assumption of independent inverse Gamma hyperpriors

$$\sigma_\epsilon^2 \sim \mathrm{IG}(a_1, b_1), \quad \sigma_\eta^2 \sim \mathrm{IG}(a_2, b_2) \tag{9}$$

for the variances σ_ϵ^2, σ_η^2. By appropriate choice of a_1, b_1, a_2, a_2, these hyperpriors are made highly dispersed.

Although we will focus on the locally adaptive model (6)-(9), some extensions are immediate: First, we may generalize the observation model (6) and the state equation (7) to the standard form $y_t = z_t' \alpha_t + \epsilon_t$, $\alpha_t = F_t \alpha_{t-1} + u_t$ of linear Gaussian state space models. Assuming again varying variances for u_t and appropriate hyperpriors as in (8), we obtain *locally adaptive state space models.* Secondly, *unequally spaced observations* can be dealt with by adjusting the variances $q_t^2 = \exp(h_t)$. For example in the case of a first order random walk, the necessary modification is $q_t^2 = \Delta_t \exp(h_t)$, where Δ_t is the difference between the t-th and the $(t+1)$-th design point.

3 Nonparametric Bayesian inference via MCMC

Fully Bayesian inference is based on the posterior $\mathrm{p}(\boldsymbol{\alpha}, \sigma_\epsilon^2, \sigma_\eta^2 \mid \mathbf{y})$ of the unknowns given the data. Due to the hierarchical structure of the model its unnormalized form is easy derived. Considering first order random walks for α_t and h_t in (7) and (8) we obtain

$$\begin{aligned}\mathrm{p}(\boldsymbol{\alpha}, \sigma_\epsilon^2, \sigma_\eta^2 \mid \mathbf{y}) &\propto \prod_{t=1}^{T} \mathrm{p}(y_t \mid \alpha_t, \sigma_\epsilon^2) \cdot \mathrm{p}(\boldsymbol{\alpha} \mid \mathbf{h}) \cdot \mathrm{p}(\mathbf{h} \mid \sigma_\eta^2) \cdot \mathrm{p}(\sigma_\epsilon^2) \cdot \mathrm{p}(\sigma_\eta^2) \\ &\propto \prod_{t=1}^{T} \mathrm{p}(y_t \mid \alpha_t, \sigma_\epsilon^2) \cdot \prod_{t=2}^{T} \mathrm{p}(\alpha_t \mid \alpha_{t-1}, h_t) \cdot \prod_{t=2}^{T} \mathrm{p}(h_t \mid h_{t-1}, \sigma_\eta^2) \\ &\quad \cdot \mathrm{p}(\sigma_\epsilon^2) \cdot \mathrm{p}(\sigma_\eta^2). \end{aligned} \tag{10}$$

Corresponding posteriors of other combinations of random walks are obtained analogously.

To sample from the posterior we use a hybrid MCMC algorithm. The required full conditionals can be derived from (10). For $\alpha_t \mid \boldsymbol{\alpha}_{s \neq t}, \mathbf{h}, \sigma_\epsilon^2, \mathbf{y}$, $t = 1, \ldots T$, we get:

$$\mathrm{p}(\alpha_t \mid \boldsymbol{\alpha}_{s \neq t}, h_s, \sigma_\epsilon^2, \mathbf{y}) \propto \mathrm{p}(y_t \mid \alpha_t, \sigma_\epsilon^2) \cdot \mathrm{p}(\alpha_t \mid \boldsymbol{\alpha}_{s \neq t}, \mathbf{h})$$

Both factors are normal distributions, namely $y_t \mid \alpha_t, \sigma_\epsilon^2 \sim \mathrm{N}(\alpha_t, \sigma_\epsilon^2)$ and $\alpha_t \mid \boldsymbol{\alpha}_{s \neq t}, \mathbf{h} \sim \mathrm{N}(\mu_t, \sigma_t^2)$ with parameters

$$\begin{aligned} \text{with} \quad \sigma_t^2 &= \begin{cases} q_{t+1}^2, & t = 1 \\ \left(1/q_t^2 + 1/q_{t+1}^2\right)^{-1}, & t = 2, \ldots, T-1 \\ q_t^2, & t = T \end{cases} \\ \text{and} \quad \mu_t &= \begin{cases} \alpha_{t+1}, & t = 1 \\ 1/\sigma_t^2 \cdot \left(q_t^2 \alpha_{t+1} + q_{t+1}^2 \alpha_{t-1}\right), & t = 2, \ldots, T-1 \\ \alpha_{t-1}, & t = T \end{cases} \end{aligned} \tag{11}$$

if α_t follows first order random walk. For a second order random walk the parameters μ_t and σ_t^2 are given by

$$\sigma_t^2 = \begin{cases} q_{t+1}^2, & t = 1 \\ \left(\frac{4}{q_{t+1}^2} + \frac{1}{q_{t+2}^2}\right)^{-1}, & t = 2 \\ \left(\frac{1}{q_t^2} + \frac{4}{q_{t+1}^2} + \frac{1}{q_{t+2}^2}\right)^{-1}, & t = 3, \ldots, T-2 \\ \left(\frac{4}{q_{t+1}^2} + \frac{1}{q_t^2}\right)^{-1}, & t = T-1 \\ q_t^2, & t = T \end{cases} \tag{12}$$

$$\text{and} \quad \mu_t \;=\; \begin{cases} 2\alpha_{t+1} - \alpha_{t+2}, & t = 1 \\ \frac{1}{\sigma_t^2} \cdot \left(\frac{2\alpha_{t+1} - \alpha_{t+2}}{q_{t+2}^2} + 2\frac{\alpha_{t+1} + \alpha_{t-1}}{q_{t+1}^2} \right), & t = 2 \\ \frac{1}{\sigma_t^2} \cdot \left(\frac{2\alpha_{t+1} - \alpha_{t+2}}{q_{t+2}^2} + 2\frac{\alpha_{t+1} + \alpha_{t-1}}{q_{t+1}^2} + \frac{2\alpha_{t-1} - \alpha_{t-2}}{q_t^2} \right), & t = 3, \ldots, T-2 \\ \frac{1}{\sigma_t^2} \cdot \left(2\frac{\alpha_{t+1} + \alpha_{t-1}}{q_{t+1}^2} + \frac{2\alpha_{t-1} - \alpha_{t-2}}{q_t^2} \right), & t = T-1 \\ 2\alpha_{t-1} - \alpha_{t-2}, & t = T. \end{cases}$$

To sample from $\alpha_t \mid \boldsymbol{\alpha}_{s \neq t}, \mathbf{h}, \sigma_\epsilon^2, \mathbf{y}$ we use the Gibbs sampler. Successive elements of the Markov chain for α_t are drawn from

$$\alpha_t \mid \boldsymbol{\alpha}_{s \neq t}, \mathbf{h}, \sigma_\epsilon^2, \mathbf{y} \sim \mathrm{N}\left(y_t + \mu_t, (\frac{1}{\sigma_\epsilon^2} + \frac{1}{\sigma_t^2})^{-1} \right).$$

In our implementation, these single move updating behaved quite well. However, other MCMC updating schemes like the block moves of Carter and Kohn (1994) and Fruehwirth-Schnatter (1994) or the MH–algorithm with conditional prior proposals of Knorr–Held (1998) are surely useful alternatives.

We use the conditional prior proposal approach for drawing from the full conditionals for $h_t, t = 2, \ldots, T$. They are factorized by $\mathrm{p}(h_t \mid \mathbf{h}_{s \neq t}, \boldsymbol{\alpha}, \sigma_\eta^2) \propto \mathrm{p}(\alpha_t \mid \boldsymbol{\alpha}_{s \prec t}, h_t) \cdot \mathrm{p}(h_t \mid \mathbf{h}_{s \neq t}, \sigma_\eta^2)$. MH–steps consists in drawing a proposal $h_t^\star$ from the conditional prior $\mathrm{p}(h_t \mid \mathbf{h}_{s \neq t}, \sigma_\eta^2)$ and accepting it with probability

$$\min \left\{ 1, \frac{\mathrm{p}(\alpha_t \mid \boldsymbol{\alpha}_{s \prec t}, h_t^\star)}{\mathrm{p}(\alpha_t \mid \boldsymbol{\alpha}_{s \prec t}, h_t)} \right\}.$$

The distribution $\mathrm{p}(\alpha_t \mid \boldsymbol{\alpha}_{s \neq t}, h_t)$ is $\mathrm{N}(\alpha_{t-1}, \exp(h_t))$ if α follows an first order random walk, respectively $\mathrm{N}(2\alpha_{t-1} - \alpha_{t-2}, \exp(h_t))$ for a second order random walk. For $\mathrm{p}(h_t \mid \mathbf{h}_{s \neq t}, \sigma_\eta^2)$ we have

$$h_t \mid \mathbf{h}_{s \neq t}, \sigma_\eta^2 \quad \propto \quad \begin{cases} \mathrm{N}(h_{t+1}, \sigma_\eta^2), & t = 1 \\ \mathrm{N}(\frac{1}{2}h_{t-1} + \frac{1}{2}h_{t+1}, \sigma_\eta^2/2), & t = 2, \ldots, T-1 \\ \mathrm{N}(h_{t-1}, \sigma_\eta^2), & t = T \end{cases}$$

if h_t follows a first order random walk. For a second order random walk it is (Knorr-Held, 1997):

$$h_t \mid \mathbf{h}_{s \neq t}, \sigma_\eta^2 \quad \propto \quad \begin{cases} \mathrm{N}(2h_{t+1} - h_{t+2},\, \sigma_\eta^2), & t = 1 \\ \mathrm{N}(\frac{2}{5}h_{t-1} + \frac{4}{5}h_{t+1} - \frac{1}{5}h_{t+2},\, \sigma_\eta^2/5), & t = 2 \\ \mathrm{N}(-\frac{1}{6}\alpha_{t-2} + \frac{2}{3}\alpha_{t-1} + \frac{2}{3}\alpha_{t+1} - \frac{1}{6}\alpha_{t-2},\, \sigma_\eta^2/6), & t = 3, \ldots, T-2 \\ \mathrm{N}(-\frac{1}{5}h_{t-2} + \frac{4}{5}h_{t-1} + \frac{2}{5}h_{t+2},\, \sigma_\eta^2/5), & t = T-1 \\ \mathrm{N}(-h_{t-2} + 2h_{t+1},\, \sigma_\eta^2), & t = T. \end{cases}$$

For the prior distributions of the two variances σ_ϵ^2 and σ_η^2 we use highly dispersed inverse Gamma distributions $\mathrm{IG}(a_i, b_i)$, $i = 1, 2$ with the parameters $a_1 = a_2 = 1$

and $b_1 = b_2 = 0.005$. With $\mathrm{p}(\sigma_\epsilon^2 \mid \mathbf{y}, \boldsymbol{\alpha}) \propto \prod_{t=1}^{T} \mathrm{p}(y_t \mid \alpha_t, \sigma_\epsilon^2) \cdot \mathrm{p}(\sigma_\epsilon^2)$ the distribution of $\sigma_\epsilon^2 \mid \mathbf{y}, \boldsymbol{\alpha}$ is also an inverse Gamma:

$$\sigma_\epsilon^2 \mid \mathbf{y}, \boldsymbol{\alpha} \sim \mathrm{IG}\left(a_1 + \frac{T}{2},\, b_1 + \frac{1}{2}\sum_{t=1}^{T}(y_t - \alpha_t)^2\right).$$

If h_t follows a first order random walk the full conditional of σ_η^2 is derived from $\mathrm{p}(\sigma_\eta^2 \mid \mathbf{h}) \propto \prod_{t=2}^{T} \mathrm{p}(h_t \mid h_{t-1}, \sigma_\eta^2) \cdot \mathrm{p}(\sigma_\eta^2)$. It is an inverse Gamma distribution, too:

$$\sigma_\eta^2 \mid \mathbf{h} \sim \mathrm{IG}\left(a_2 + \frac{T}{2},\, b_2 + \frac{1}{2}\sum_{t=2}^{T}(h_t - h_{t-1})^2\right).$$

In the case of a second order random walk the full conditional is obtained analogously:

$$\sigma_\eta^2 \mid \mathbf{h} \sim \mathrm{IG}\left(a_2 + \frac{T}{2},\, b_2 + \frac{1}{2}\sum_{t=3}^{T}(h_t - 2h_{t-1} - h_{t-2})^2\right).$$

4 Examples

To gain experience with practical performance our locally adaptive approach was applied to a number of simulated and real data. Subsection 4.1 deals with jumps in simple step functions, Subsection 4.2 reports on the results for the unsmooth functions constructed for wavelet shrinkage by Donoho and Johnstone (1994), and Section 4.3 contains an application to a time series of sales data from West and Harrison (1989).

4.1 Jumps

We first study empirically the approach for noisy observations of a simple step function $\boldsymbol{\alpha}$ with a jump to a higher level and a further jump back to the original level. Data were generated at 150 design points according to (1), with jumps at $t = 51$ and $t = 101$. The first row of Figure 4.1 shows three step functions together with the generated data. The height of the jumps and the variance of the observations is different, with the jump to noise ratio increasing from the left to the right. The aim of this and simular studies was to investigate the impact of the jump to noise ratio and the different smoothness priors for $\boldsymbol{\alpha}$ and $\mathbf{h}$ on the quality of fitting step functions and recognizing jumps or change points.

The estimates of $\boldsymbol{\alpha}$ and $\mathbf{q}$ in the second and the third row of Figure 4.1 are obtained by choosing a RW(1) for $\boldsymbol{\alpha}$ and $\mathbf{h}$. The estimated variance function increases exactly at the jumps if the jump to noise ratio is high enough as in the right column. For the other two situations, local adaption is not satisfactory. However, such a low jump to noise ratio will also pose problems for other nonparametric function estimators.

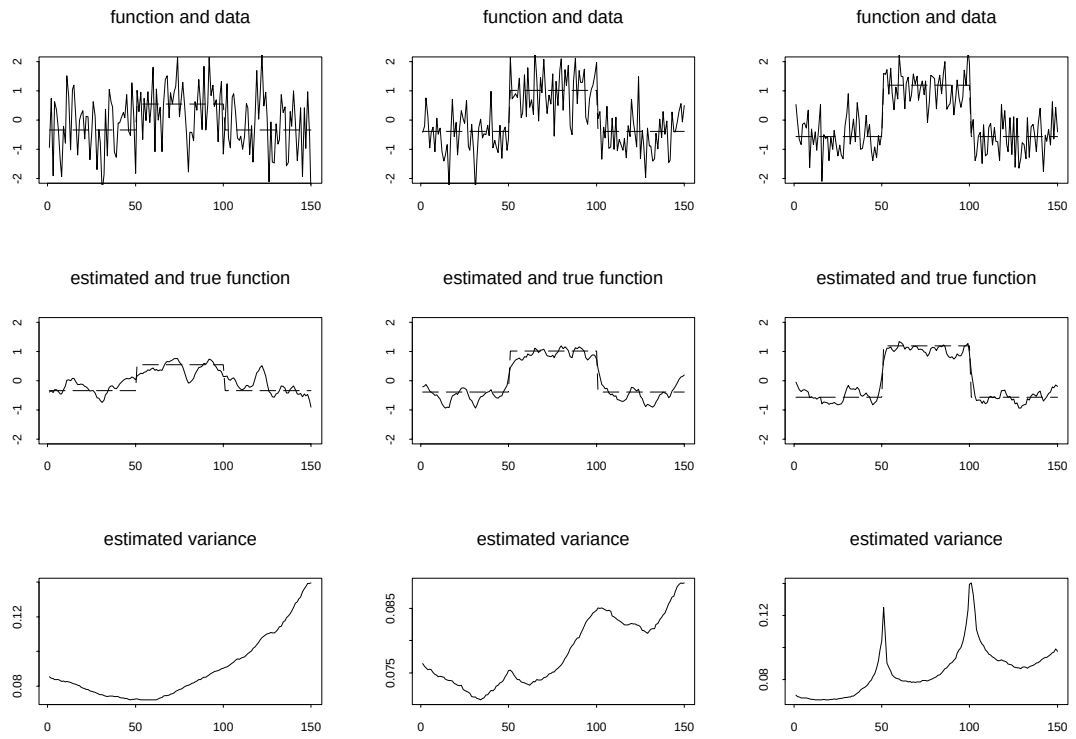

Figure 4.1: Jump functions are becoming more distinct from the left to the right. The first row shows the true function and its noisy version, the second one the estimated function compared to the real one and the third one the estimated variances $\mathbf{q}$. Both $\boldsymbol{\alpha}$ and $\mathbf{h}$ are modelled by a RW(1).

Next we considered fitting behaviour for different combinations of first and second order random walks as smoothness prior for $\boldsymbol{\alpha}$ and $\mathbf{h}$. As to be expected, Figures 4.2 and 4.3 clearly indicate that RW(1) smoothness priors for the variance function are most suitable for fitting step functions. In particular in combination with a RW(1) for $\boldsymbol{\alpha}$, jumps are well detected by corresponding peaks in the estimated variance function $\mathbf{q}$. Second order random walks for $\mathbf{h}$ are far less suitable here, since they lead to a rather oscillating estimate $\hat{\boldsymbol{\alpha}}$ in the constant parts of the step function.

4.2 Blocks, Bumps, Doppler and Heavy Sine

We now demonstrate performance for the four datasets Blocks, Bumps, Doppler and Heavy Sine constructed for wavelet shrinkage by Donoho and Johnstone (1994). Each of the datasets, shown in Figure 4.4, consists of 2048 observations and has a signal to noise ratio of 7.

Let us first take a closer look at the dataset Blocks in Figure 4.5. The best and also very satisfying estimate for $\boldsymbol{\alpha}$ is again obtained with a first order random walk for $\boldsymbol{\alpha}$ and $\mathbf{h}$. Choosing a second order random walk for $\mathbf{h}$ but keeping a first order random walk for $\boldsymbol{\alpha}$ still yields a good fit. In both cases the estimated variances $\hat{\mathbf{q}}$ in Figure 4.6 have distinctly high values exactly at the design points with jumps. Choice of a second order random walk for $\mathbf{h}$ however leads to the same oscillating

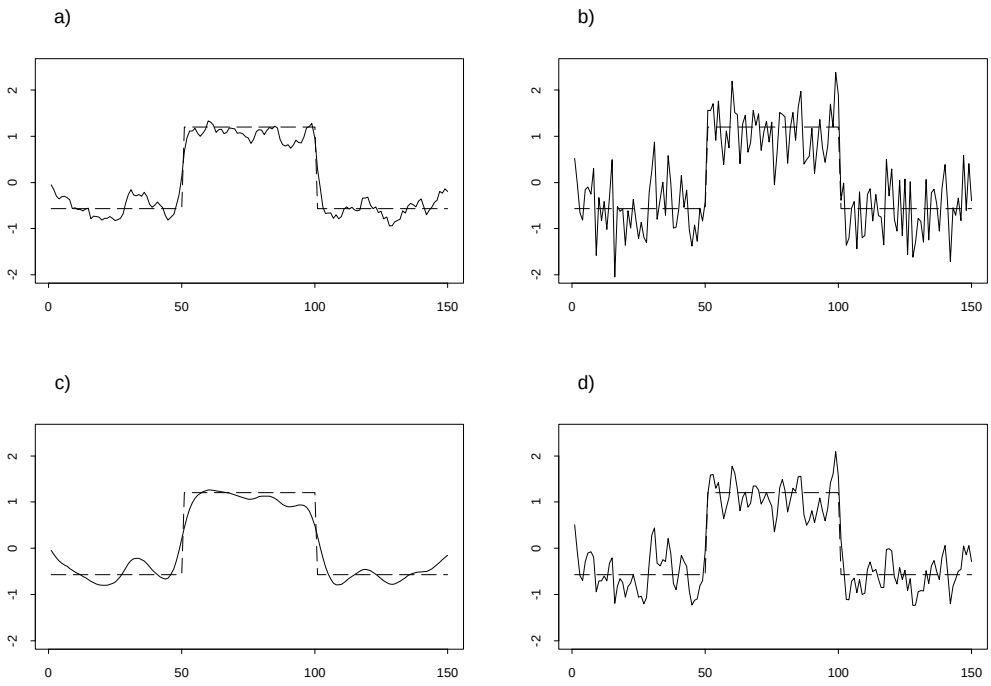

Figure 4.2: Estimates $\hat{\boldsymbol{\alpha}}$ for different combinations of the random walks for $\boldsymbol{\alpha}$ and $\mathbf{h}$ with **a)** both RW(1), **b)** $\boldsymbol{\alpha}$ RW(2), $\mathbf{h}$ RW(1), **c)** $\boldsymbol{\alpha}$ RW(1), $\mathbf{h}$ RW(2) and **d)** both RW(2).

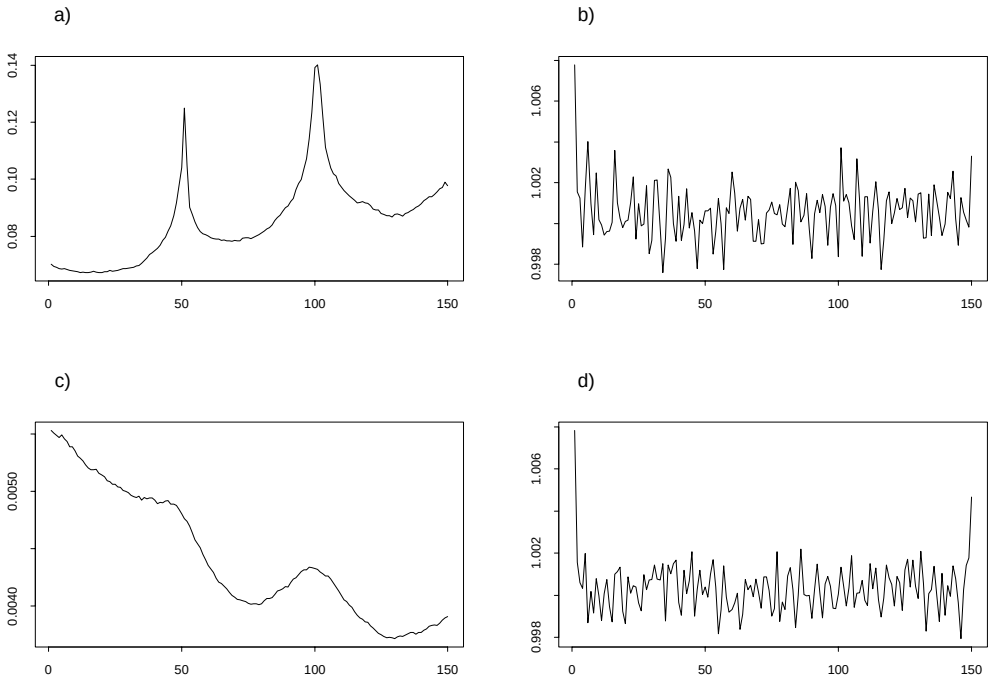

Figure 4.3: Estimated variance function $\hat{\mathbf{q}}$ for different combinations of the random walks for $\boldsymbol{\alpha}$ and $\mathbf{h}$ with **a)** both RW(1), **b)** $\boldsymbol{\alpha}$ RW(2), $\mathbf{h}$ RW(1), **c)** $\boldsymbol{\alpha}$ RW(1), $\mathbf{h}$ RW(2) and **d)** both RW(2).

behaviour as in Subsection 4.1 (Figure 4.2, right column). Choice of a second order random walk for both $\boldsymbol{\alpha}$ and $\mathbf{h}$ leads to a variance estimate that is nearly constant about 1.

For better resolution, Figure 4.7 displays only parts of the data and their fitted values for each of the four datasets considered. Corresponding estimates of the variances $\mathbf{q}$ are given in Figure 4.8. As already seen for the dataset Blocks, bumps are well reproduced while more constant parts have been smoothed. The ability of our locally adaptive model to react appropriately to situations where a higher variance is needed is absolutely convincing for the dataset Doppler. Here the sine curve is oscillating more and more when it is coming closer to zero. Obviously an increasing variance is required to deal with this situation, and exactly this behaviour is observed for the estimated variance $\mathbf{q}$. For the dataset Heavy Sine the results are not so convincing. The sudden jump near $t = 600$ is well detected, but in the other parts with only a moderate curvature the fit is too rough.

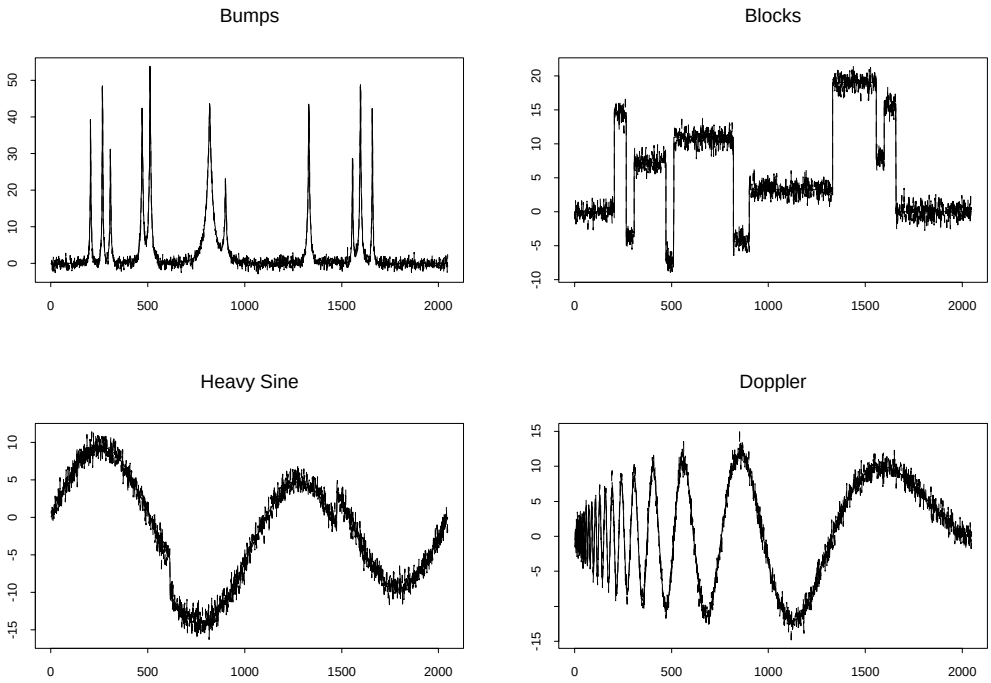

Figure 4.4: Datasets Bumps, Blocks, Heavy Sine and Doppler.

4.3 CP6 sales data

The monthly CP6 sales data (West and Harrison, 1989) shown on the left in Figure 4.9 indicate an additive outlier and a change of the slope in December 1955 as well as further change points in January 1957 and 1958. The best fit for the trend $\boldsymbol{\alpha}$ displayed on the right in Figure 4.9 was obtained here by a first order random walk for $\boldsymbol{\alpha}$ and a second order random walk for $\mathbf{h}$. Adaption to the change points and to smooth trends between them seems to be quite adequate. Also the changepoints are clearly detected by the distinct peaks in the fit of the variance function.

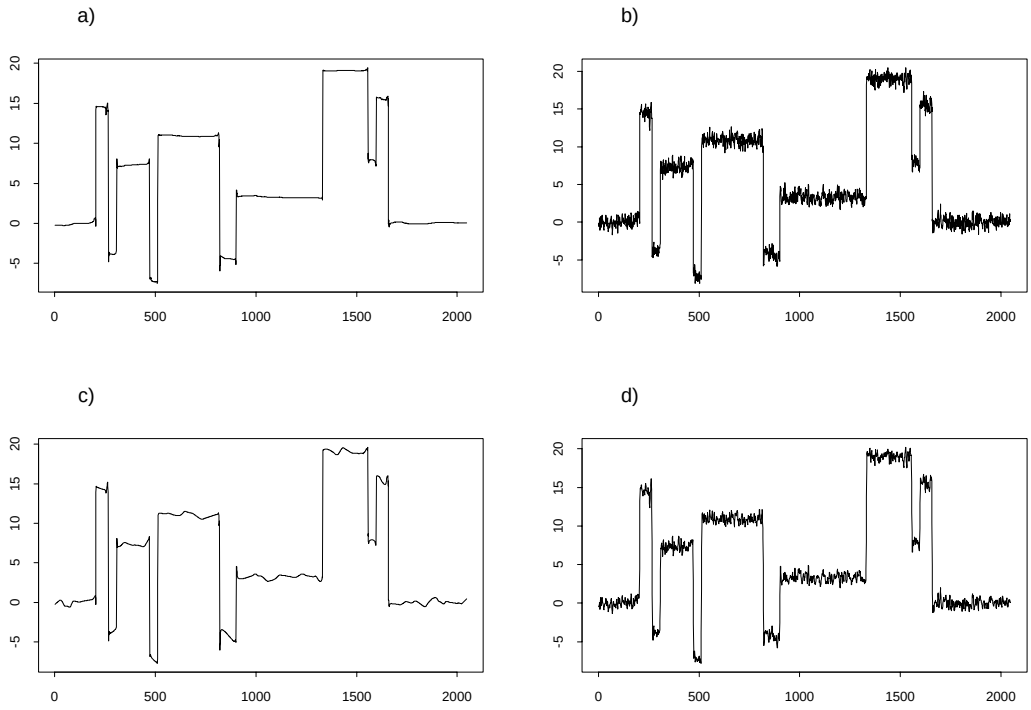

Figure 4.5: Estimates $\hat{\boldsymbol{\alpha}}$ of Blocks for different combinations of the random walks for $\boldsymbol{\alpha}$ and **h** with **a)** both RW(1), **b)** $\boldsymbol{\alpha}$ RW(2), **h** RW(1), **c)** $\boldsymbol{\alpha}$ RW(1), **h** RW(2) and **d)** both RW(2).

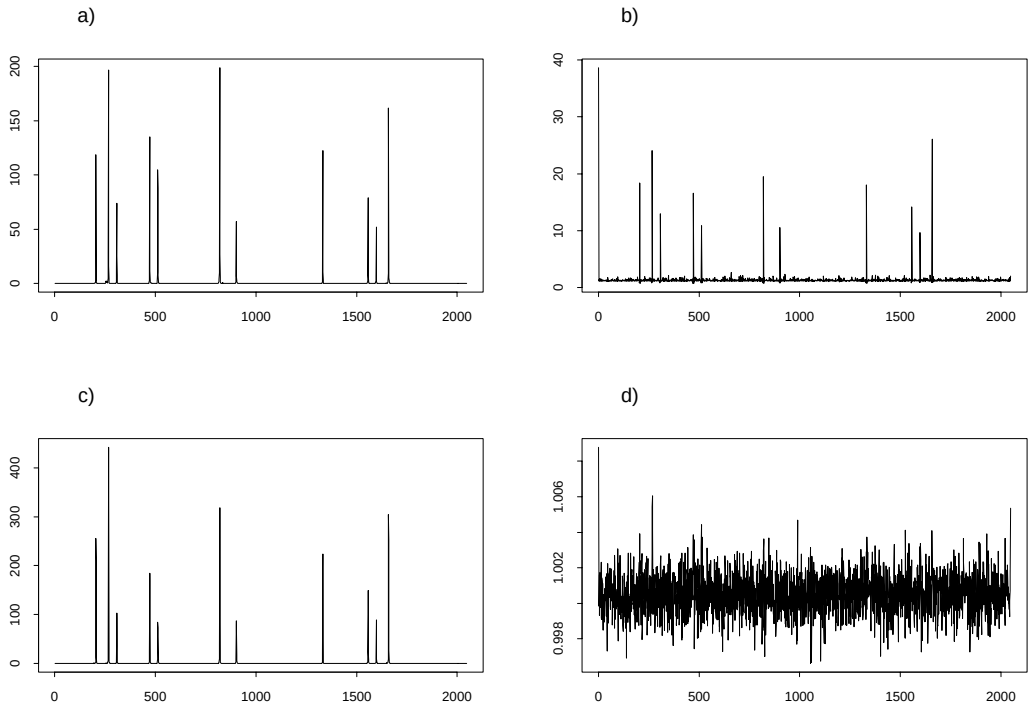

Figure 4.6: Estimated variance function $\hat{\mathbf{q}}$ of Blocks for different combinations of the random walks for $\boldsymbol{\alpha}$ and **h** with **a)** both RW(1), **b)** $\boldsymbol{\alpha}$ RW(2), **h** RW(1), **c)** $\boldsymbol{\alpha}$ RW(1), **h** RW(2) and **d)** both RW(2).

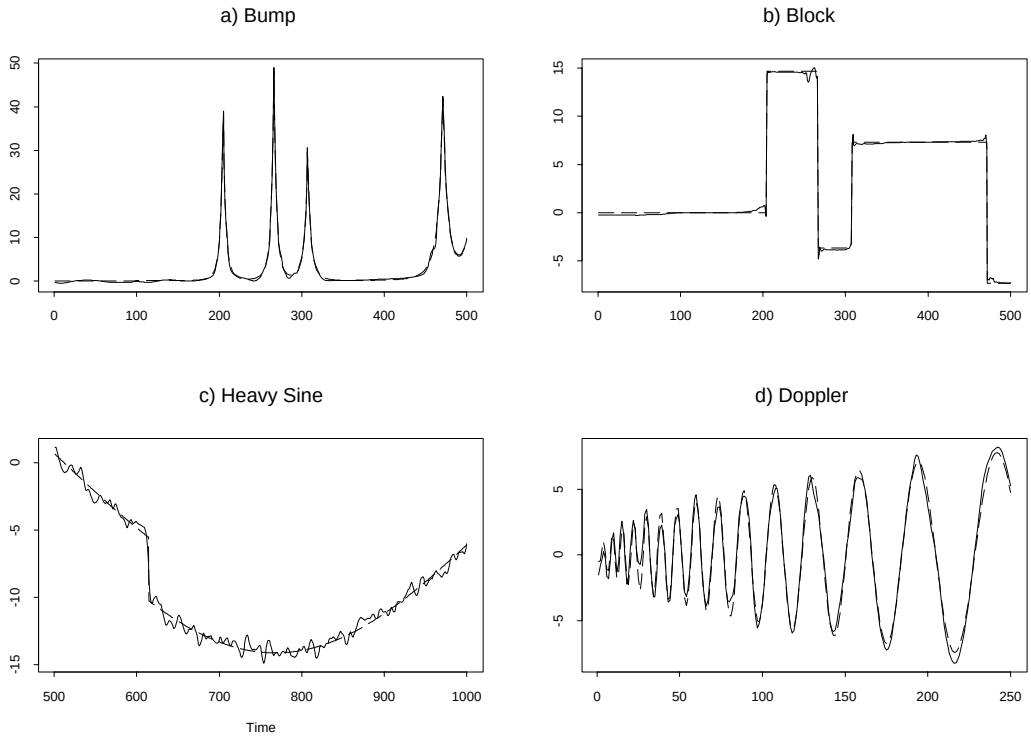

Figure 4.7: Parts of fitted functions for the datasets in Figure 4.4 with the following random walk combinations: **b)** both RW(1) and **a)**, **c)** and **d)** $\boldsymbol{\alpha}$ RW(2), **h** RW(1).

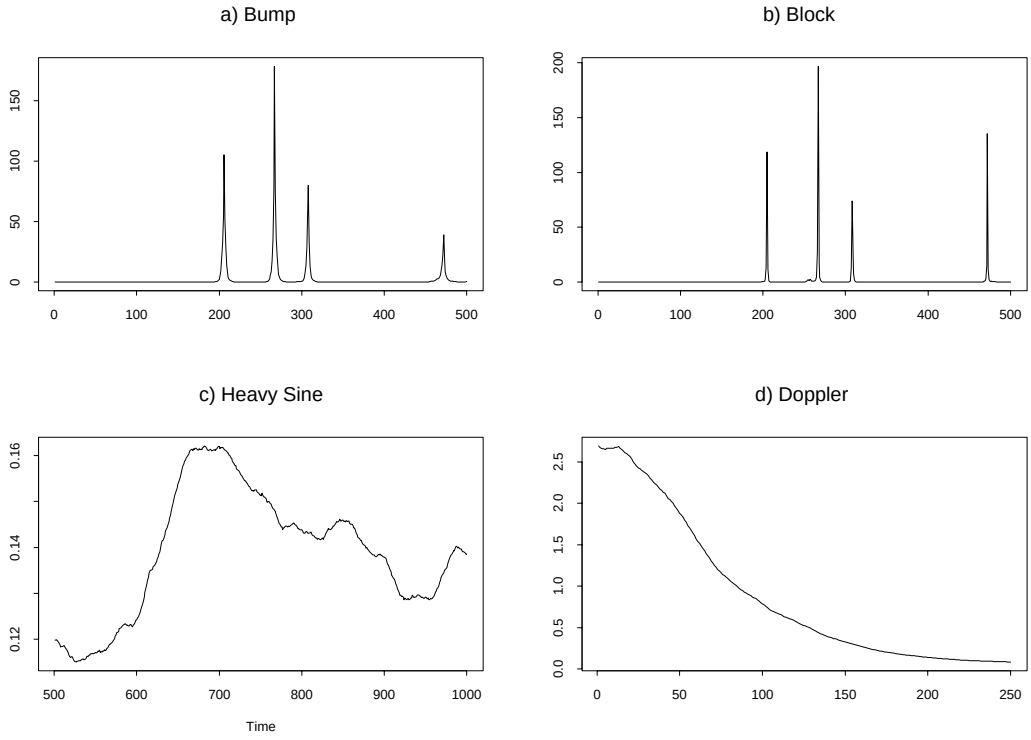

Figure 4.8: Estimated variance function $\hat{\mathbf{q}}$ of the parts in Figure 4.4 with the following random walk combinations: **b)** both RW(1) and **a)**, **c)** and **d)** $\boldsymbol{\alpha}$ RW(2), **h** RW(1).

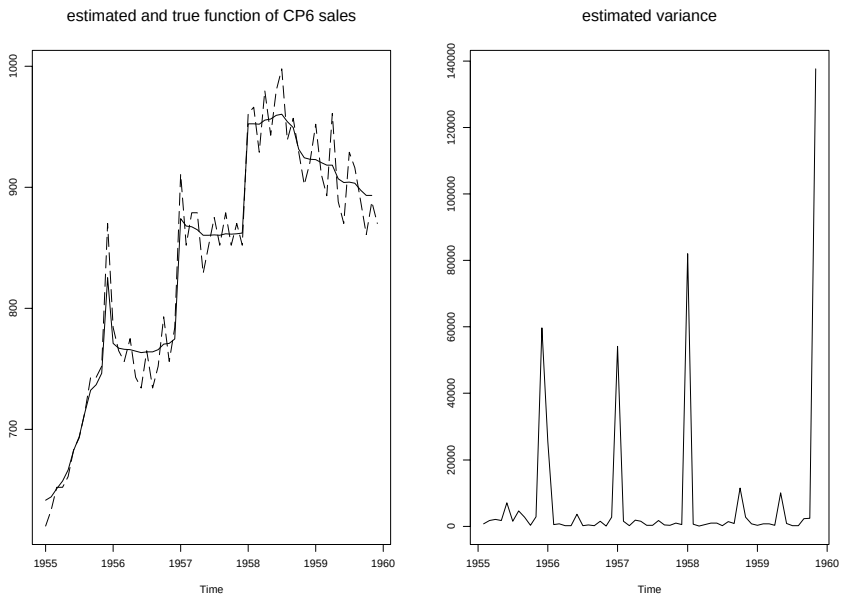

Figure 4.9: On the left CP6 sales data and the best estimate of $\boldsymbol{\alpha}$ which was achieved by a RW(1) for $\boldsymbol{\alpha}$ and a RW(2) for $\mathbf{h}$. The corresponding estimated variances $\mathbf{q}$ are given on the right.

5 Conclusions

The results in Section 4 provide empirical evidence that locally adaptive dynamic models are a promising and conceptually simple approach for nonparametric estimation of unsmooth curves. In particular, the results for blocks, bumps and the Doppler curve are very encouraging.

Obviously, some data driven method for model choice, in particular giving support for deciding about the types of random walks, would be rather helpful. The recently proposed DIC criterion (Spiegelhalter et al., 1998) is a rather general tool in connection with MCMC techniques and we intend to adapt it to our specific situation.

Apart from the extensions already mentioned at the end of Section 2, the following generalizations could offer a field for future research: First, the Gaussian observation model (6) can be replaced by non Gaussian observation models. In particular, choice of distribution from the exponential family defines a large class of locally adaptive modifications of standard dynamic generalized linear models (e.g. Fahrmeir and Tutz,1994, ch.8). A further possibility is the introduction of varying variances in the observation model, as in stochastic volatility models (Taylor, 1986). Another generalization concerns Markov random fields for spacial data analysis. Here local adaption for unsmooth surfaces could be achieved by introducing unknown weights or scale factors in pairwise difference priors (see e.g. Besag, Green, Higdon and Mengersen, 1995, Section 3), together with spatial smoothness priors for them, and estimate these weights simultaneously with the surfaces.

Acknowledgement: This research has been supported by grants from the German National Sience Foundation, Sonderforschungsbereich 386. We also thank Leo Knorr–Held and Stefan Lang for useful discussions.

References

Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995). *Bayesian Computation and Stochastic Systems.* Statistical Science 10, 3-66.

Carter, C. K. and Kohn, R. (1994). *On Gibbs sampling for state space models.* Biometrica 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). *Multivariate Statistical Modelling Based on Generalized Linear Models.* Springer-Verlag. New York.

Fan, J. and Gijbels, I. (1995). *Data–driven bandwidth selection in local polynomial filting: variable bandwidth and spatial adaption.* Journal of the Royal Statistic Society B57, 371-394.

Fruehwirth-Schnatter, S. (1994). *Data augmention 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 University Press. Cambridge.

Knorr–Held, L. (1997). *Hierarchical Modelling of Discrete Longitudinal Data.* Herbert Utz Verlag. München.

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

Spiegelhalter, D. J., Best, N. G., and Carlin, B. P. (1998). *Bayesian deviance, the effective number of parameters, and the comparison of arbitrary complex models.* Preprint.

Taylor, St. J. (1986). *Modelling Financial Time Series.* John Wiley, Chichester.

West, M. and Harrison, P. J. (1989). *Bayesian Forecasting and Dynamic Models.* New York, Springer.

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