



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

INSTITUT FÜR STATISTIK
SONDERFORSCHUNGSBEREICH 386



Fahrmeir:

Recent Advances in Semiparametric Bayesian Function Estimation

Sonderforschungsbereich 386, Paper 137 (1998)

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

Projektpartner



Recent Advances in Semiparametric Bayesian Function Estimation

Ludwig Fahrmeir
Institut für Statistik, Universität München
Ludwigstr. 33
80539 München
e-mail: fahrmeir@stat.uni-muenchen.de

Abstract: Common nonparametric curve fitting methods such as spline smoothing, local polynomial regression and basis function approaches are now well developed and widely applied. More recently, Bayesian function estimation has become a useful supplementary or alternative tool for practical data analysis, mainly due to breakthroughs in computerintensive inference via Markov chain Monte Carlo simulation. This paper surveys recent developments in semiparametric Bayesian inference for generalized regression and outlines some directions in current research.

1 Introduction

Regression analysis is one of the most widely used methods in applied statistics. Often it is difficult to prespecify parametric models, and nonparametric fitting of unknown regression functions is needed. Common methods are kernel-based regression, spline smoothing, local polynomial regression and basis function approaches such as regression splines, Fourier expansions and wavelets. Very recently, semiparametric Bayesian methods have been developed and are a promising alternative tool for practical data analysis. They rely on Markov chain Monte Carlo (MCMC) simulation and provide rich output for inference. No approximate normality conjectures for estimators are required, so that the methods are also useful for moderate sample sizes and in complex, high-dimensional problems. This paper surveys recent advances in Bayesian function estimation, distinguishing smoothing priors and basis function approaches as two mainstream directions. We do not give details about MCMC algorithms here, but refer to general introductions to Bayesian data analysis, as well as to original work mentioned in later sections. Section 2 deals with models for Gaussian response variables and Section 3 with the non-Gaussian case. In particular for fundamentally non-Gaussian response such as binary or other discrete responses, there is clear need for additional research. Section 4 points out some extensions and ideas for future developments.

2 Gaussian nonparametric regression

For bivariate observations $(y_1, x_1), \dots, (y_n, x_n)$ the classical nonparametric regression model on a response variable Y and a dependent variable X is

$$y_i = f(x_i) + \varepsilon_i, \quad i = 1, \dots, n, \quad (1)$$

with i.i.d Gaussian errors $\varepsilon_i \sim N(0, \sigma^2)$, and an unknown regression function f that we want to estimate. In the following, $y = (y_1, \dots, y_n)^T$ is the vector of observations on Y , and we will not distinguish notationally between the function f and the vector $f = (f(x_1), \dots, f(x_n))^T$ of function evaluations. This section describes and outlines some recent Bayesian approaches for estimating f through posterior sampling by Markov chain Monte Carlo (MCMC) techniques. Roughly, we may distinguish between methods based on roughness penalties or smoothness priors and methods based on basis functions.

One of the attractive features of these methods is that they are tailor-made for extension to the case of multivariate regressors X_1, \dots, X_p by using additive models and, somewhat more generally, semiparametric additive or varying coefficient models.

2.1 The smoothness prior approach for univariate regression

For simplicity first consider the case of equidistant design points or observations x_1, \dots, x_n . For nonparametric estimation based on smoothness priors, the observation model (1) is supplemented by assigning an appropriate prior to f . Common local smoothness priors are random walks of first (RW(1)) and second (RW(2)) order

$$f(x_i) - f(x_{i-1}) = u_i \quad \text{or} \quad f(x_i) - 2f(x_{i-1}) + f(x_{i-2}) = u_i, \quad (2)$$

with i.i.d. Gaussian errors $u_i \sim N(0, \tau^2)$. Initial values are specified by $f(x_1) \sim N(0, c)$ or $f(x_1), f(x_2) \sim N(0, cI)$, where we usually choose diffuse priors, corresponding to the limiting case $c \rightarrow \infty$. The difference equations (2) penalize deviations from locally constant functions $f(x_i) = f(x_{i-1})$ or straight lines $f(x_i) = 2f(x_{i-1}) - f(x_{i-2})$. The penalty increases as the variance τ^2 becomes smaller. The random walk models (2) define a multivariate Gaussian smoothness prior $p(f)$ for f . Also, the observation model (1) defines a multivariate Gaussian $p(y | f) \sim N(f, \sigma^2 I)$ for data y given f . For the moment, the variances or hyperparameters σ^2 and τ^2 are regarded as known or given constants. Thus, the posterior $p(f | y) \propto p(y | f) p(f)$ is also Gaussian and characterized by the posterior mean $\hat{f} = E(f | y)$ and covariance $\text{Var}(f | y)$. Due to normality, \hat{f} is also the posterior mode. Assuming diffuse initial priors, \hat{f} can therefore also be obtained as a minimizer of the

penalized least squares criterion

$$\sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \sum_{i=d+1}^n (\nabla^d f(x_i))^2 \longrightarrow \min_f \quad (3)$$

with ∇^d , $d = 1, 2$, as the first or second order difference operator and smoothness parameter $\lambda = \sigma^2 / \tau^2$. Estimation of f via (3) is the classical “method of graduation” of Whittaker (1923). Obviously, (3) has a non-Bayesian interpretation, with the first term as a measure of goodness of fit, the second as a roughness penalty and λ controlling the bias - variance trade off. The penalty terms are discretized versions of corresponding penalty terms

$$\int [f'(x)]^2 dx \quad \text{resp.} \quad \int [f''(x)]^2 dx, \quad (4)$$

leading to quadratic resp. cubic spline smoothing. Already for a moderate number of observations, smoothing splines and discretized versions \hat{f} are virtually indistinguishable. Therefore, solutions of (3) are sometimes called discrete splines. In matrix notation, (3) can be written as

$$(y - f)^T (y - f) + \lambda f^T K f \longrightarrow \min_f \quad (5)$$

with appropriately defined penalty matrix K . For example, one gets

$$K = \begin{pmatrix} 1 & -1 & & & & \\ -1 & 2 & -1 & & & 0 \\ & \ddots & \ddots & \ddots & & \\ 0 & & -1 & 2 & -1 & \\ & & & -1 & 1 & \end{pmatrix}$$

for a first order random walk.

The solution of (5) and thus the posterior mean or mode estimate \hat{f} is given by

$$\hat{f} = E(f|y) = S y$$

with smoother matrix $S = (I + \lambda K)^{-1}$. Within the Bayesian framework, we are interested not only in the posterior mean but in the entire posterior distribution $p(f|y)$. The observation model (1) is $y|f \sim N(0, \sigma^2 I)$, from (2) one obtains the global smoothness prior

$$f \sim N(0, K^{-1} \tau^2), \quad (6)$$

where K^{-} is a generalized inverse of K , and the posterior is

$$f|y \sim N(S y, S \sigma^2). \quad (7)$$

Since (1) and (2) define a state space model with “states” $f(x_1), \dots, f(x_n)$, the linear Kalman filter and smoother computes posterior means $f(x_i)|y$

and variances $\text{Var}(f(x_i) | y)$, i.e. the diagonal of $S \sigma^2$, in $O(n)$ operations. Moreover, and this seems to be far less known, the complete covariance matrix $S \sigma^2$ as well as a Cholesky square root $S^{1/2} \sigma^2$, can be computed easily from its output, see Fahrmeir and Kaufmann (1991, Prop. 1) for general state space models. For the special case considered here, this leads to the banded Cholesky factorization $(I + \lambda K) / \sigma^2 = L L^T$, so that $S^{1/2} \sigma = L^{-1}$ can be computed in $O(n)$ operations. Therefore, a posterior realization of f can be obtained from

$$f = S y + S^{1/2} \sigma z = \hat{f} + S^{1/2} \sigma z, \quad z \sim N(0, I).$$

Of course, the banded Cholesky decomposition of $(I + \lambda K)$ can also be efficiently computed without making direct use of the Kalman filter and smoother, as suggested by Hastie and Tibshirani (1998) for cubic smoothing splines. Other suggestions for efficient posterior sampling of f have been made by Fruewirth - Schnatter (1994), Carter and Kohn (1994) and de Jong and Shephard (1996) in the context of linear state space models. It may appear that posterior sampling from Gaussian models is not very interesting, since the posterior is Gaussian and its first and second moments can be computed already with efficient algorithms. The point is, however, that these methods are important building blocks for posterior sampling in more complex situations.

Up to now smoothing parameters or variances σ^2 and τ^2 were considered fixed or known. In practice they have to be determined as well. In a full Bayesian approach priors are put on the variances and their posteriors are estimated along with the function f . The standard choice are independent highly dispersed inverse gamma priors $p(\sigma^2) \sim IG(a_0, b_0)$, $p(\tau^2) \sim IG(a_1, b_1)$. The posterior is then

$$p(f, \sigma^2, \tau^2 | y) \propto p(y | f, \sigma^2) p(f | \tau^2) p(\sigma^2) p(\tau^2).$$

Posterior sampling can be carried out with the Gibbs sampler: Realizations are drawn sequentially from $p(f | y, \sigma^2, \tau^2)$, and from $p(\sigma^2 | y, f, \tau^2) = p(\sigma^2 | y, f)$, $p(\tau^2 | y, f, \sigma^2) = p(\tau^2 | f)$, which are inverse Gaussian with updated parameters, see e.g. Fruewirth - Schnatter (1994), Carter and Kohn (1994), Hastie and Tibshirani (1998). There is empirical evidence that these “multi-move” sampling schemes are superior in convergence and mixing behaviour to “single-move” samplers that update single components $f(x_i)$ of f , given the rest (Carlin, Polson and Stoffer, 1992).

For non-equally spaced observations, smoothness priors have to be modified appropriately to account for non-equal distances $\delta_i = x_i - x_{i-1}$ between observations. RW(1) models are now specified by

$$f(x_i) = f(x_{i-1}) + u_i, \quad u_i \sim N(0; \delta_i \sigma^2),$$

i.e. by adjusting error variances from σ^2 to $\delta_i \sigma^2$. RW(2) models generalize to

$$f(x_i) = \left(1 + \frac{\delta_i}{\delta_{i-1}}\right) f(x_{i-1}) - \frac{\delta_i}{\delta_{i-1}} f(x_{i-2}) + u_i, \quad u_i \sim N(0; g_i \sigma^2),$$

where g_i is an appropriate weight. The simplest weight is again $g_i = \delta_i$, but more complicated weights can be derived from underlying continuous priors or with other arguments, see e.g. Knorr-Held (1997). A related, but different AR(2) model can be found in Berzuini and Larissa (1996). Another possibility is to work with stochastic differential smoothness priors for smoothing splines and state space models for non-equally spaced observations derived from them, see Carter and Kohn (1994), Wong and Kohn (1996). Hastie and Tibshirani (1998) directly start with smoothness priors as in (6), where K is the penalty matrix corresponding to smoothing splines.

2.2 Unsmooth functions

So far the methods are useful for estimating and exploring smooth regression functions, that means functions with curvatures that do not exhibit too much spatial inhomogeneity. They have problems, however, with functions that are more wiggly or have edges and jumps. This is a consequence of the global variances or smoothing parameters, which are assumed constant over the design space, and the Gaussian prior assumptions, leading to quadratic roughness penalties.

One approach to handle this problem is therefore to replace to Gaussian priors in (2) or (6) by heavy tail priors as they are known in robust statistics. Prominent examples are finite mixtures of normals, Student priors with low degrees of freedom, Laplace, Huber or truncated Gaussian priors. A posterior mode or penalized likelihood approach is given in Fahrmeir and Künstler (1998), and posterior sampling approaches are described in Carter and Kohn (1996). An interesting family of distributions, that contains Gaussian and Laplace densities as special cases, is the exponential power family (Box and Tiao, pp.156-243). Its potential for flexible modelling has still to be explored.

The second possibility to deal with unsmooth functions is to allow for variances that vary over the design space. This corresponds to locally varying smoothness parameters or bandwidths, which have gained considerable interest in kernel-based nonparametric regression. A Bayesian approach is proposed by Fronk and Fahrmeir (1998). They modify the local RW(1) and RW(2) priors (2) by assuming varying variances τ_i^2 for the errors u_i . Reparameterizing by $h_i = \log(\tau_i^2)$ and adding random walk smoothness priors for the sequence $h = (h_1, \dots, h_n)$ in a further layer of the hierarchy, leads to locally adaptive dynamic models with observation equation (1), random walk priors (2), but with errors

$$u_i \sim N(0; \exp(h_i)), \quad i = 1, \dots, n$$

and RW(1) or RW(2) priors

$$h_i = h_{i-1} + v_i, \quad h_i = 2h_{i-1} - h_{i-2} + v_i, \quad v_i \sim N(0; q),$$

and an inverse Gamma prior for q .

Posterior sampling is carried out by a hybrid MCMC algorithm, combining

Gibbs and Metropolis-Hastings steps. The method works quite well with the “Blocks”, “Bumps” and “Doppler” simulated examples, constructed by Donoho and Johnstone (1994) for wavelet shrinkage.

2.3 Basis function approach

Let $S = \{B_i(x); i \in I\}$ be a set of linearly independent univariate functions, which are called basis functions. This section outlines recent Bayesian methods for nonparametric estimation of f by modeling it as a linear combination

$$f(x) = \sum_{i \in I} \beta_i B_i(x) \quad (8)$$

of these basis functions. Popular choices for the B_i are various bases for spline functions, piecewise polynomials and wavelets. Generally, it is difficult to determine which basis functions should be included in (8). If the basis has too many functions, estimates for β_i will have high variability or may even produce interpolation in the extreme case. Conversely, if too few functions are in the basis, the resulting estimator will be severely biased.

Smith and Kohn (1996) propose to approximate $f(x)$ by cubic regression splines using the truncated power series basis

$$f(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3 + \sum_{k=1}^m \beta_k (x - \tau_k)_+^3,$$

where $\tau_1 < \dots < \tau_m$ is a large number of potential “knots” placed along the domain of the independent variable x , and $(z)_+ = \max(0; z)$. A Bayesian variable selection approach is used to determine the significant knots and to estimate the parameters β . Denison et al. (1998) use a wider class of piecewise polynomials as basis functions that incorporates smoothing splines as a subclass. However, it is more flexible because also unsmooth functions with rapidly varying first and second derivatives or even discontinuities can be modelled adequately. They assume

$$f(x) = \sum_{j=0}^l \alpha_j (x - \tau_0)_+^j + \sum_{m=1}^k \sum_{j=l_0}^l \beta_{jk} (x - \tau_m)_+^j,$$

where τ_0 is the left boundary knot, l is the order of the piecewise polynomials and l_0 gives the degree of continuity at the knot points. Taking $l = l_0 = 3$ gives the cubic regression spline basis as a special case. The number and location of knots is considered as unknown, and sampling from posteriors for both the number and the location of knots is addressed using the reversible jump MCMC simulation technique of Green (1995). The method is some hybrid technique between an empirical and a full Bayesian approach in that the α 's and β 's are estimated by least squares conditional upon number and location of knots as well as $\sigma^2 = \text{var}(\varepsilon_i)$. The novelty is that the resulting posterior mean estimate for f corresponds to Bayesian model averaging.

The Bayesian basis function approach is surely rather promising and it seems worthwhile to extend it to other function bases, in particular those with local support, like B-Splines, or orthogonal basis functions, like the Demmler-Reinsch basis, for smoothing splines, or to orthogonal wavelet bases.

2.4 Extensions to additive models

Without further structural assumptions, nonparametric regression with more than two regressors quickly becomes intractable due to the interactions which have to be included. Additive models for observations $(y_i, x_{i1}, \dots, x_{ip})$, $i = 1, \dots, n$, on a Gaussian response variable Y and a vector $X = (X_1, \dots, X_p)$ of metrical covariates assume

$$y_i = \alpha + \sum_{j=1}^p f_j(x_{ij}) + \varepsilon_i, \quad \varepsilon_i \sim N(0; \sigma^2). \quad (9)$$

For identifiability, some constraint has to be imposed, for example assuming that the unknown functions are centered by

$$\sum_{i=1}^n f_j(x_{ij}) = 0.$$

The standard nonparametric approach for fitting the f_j is to use univariate smoothers and the backfitting algorithm as described in Hastie and Tibshirani (1990). Extensions of additive models that can be treated with more or less the same techniques are semiparametric additive models

$$y_i = \alpha + \sum_{j=1}^p f_j(x_{ij}) + w_i' \beta + \varepsilon_i, \quad (10)$$

where w_i is a vector of further covariates whose effect is assumed to be linear, and varying coefficient models

$$y_i = \alpha + \sum_{j=1}^p f_j(x_{ij}) z_{ij} + w_i' \beta + \varepsilon_i \quad (11)$$

with additional design variables or covariates z_j . The Gibbs sampler is tailor-made for drawing posteriors from additive models and can be interpreted as ‘‘Bayesian backfitting’’ (Hastie and Tibshirani, (1998)). Suppressing dependence on hyperparameters like variances etc., the generic form of the Gibbs sampling algorithm is as follows:

- Step 1 initialize - $f_0 = \alpha = \text{ave}(y_i)$, $f_j = f_j^0$, $j = 1, \dots, p$
- Step 2 cycle - $j = 0, \dots, p, 0, \dots, p, \dots$ draw posterior samples from $p(f_j | f_1, \dots, f_{j-1}, f_{j+1}, \dots, f_p, y)$
- Step 3 continue step 2 until the joint distribution of (f_0, \dots, f_p) doesn't change.

In Step 2, one of the univariate function samplers can be used. For example, if S_j is one of the smoothing matrices corresponding to random walk or smoothing spline models, then Step 2 becomes

Step 2 (Hastie and Tibshirani, (1998)).
 iterate for $t = 1, 2, \dots$
 - do for $j = 0, 1, \dots, p$: Define partial residual
 $r_j^t = y - \sum_{k < j} f_k^t - \sum_{k > j} f_k^{t-1}$
 Generate $z_j^t \sim N(0; I)$ and update
 $f_j^t \leftarrow S_j r_j^t + \sigma S_j^{\frac{1}{2}} z_j^t$

The algorithm can easily be extended to the more general models (10) and (11).

3 Nonparametric regression for non-Gaussian responses

There are two mainstreams for regression analysis with non-Gaussian responses Y . In the first approach, one tries to find a suitable transformation $\tilde{Y} = T(Y)$ such that the transformed variable \tilde{Y} is, at least approximately, Gaussian. A well known class of such transformations are Box-Cox-type transformations. Given the transformation, regression techniques for Gaussian responses are applied. A Bayesian approach for a data driven choice of transformations is suggested in Smith and Kohn (1996). This section deals with the second approach, where the distribution of Y is directly modelled by a non-Gaussian distribution. We distinguish between so-called conditionally Gaussian models, where the density of Y is Gaussian conditional upon a further variable, often a mixture variable, and fundamentally non-Gaussian models, e.g. for regression analysis with categorical responses.

3.1 Conditionally Gaussian models

Historically, conditionally Gaussian models were first used for additive outliers. Let $k_i \in \{1, 2\}$, $i = 1, \dots, n$, denote an i.i.d. sequence of unobserved indicator variables, with $k_i = 1$ for an ‘‘usual’’ observation y_i and $k_i = 2$ for an outlier y_i . Conditional upon k_1, \dots, k_n the observation model is Gaussian,

$$y_i = f(x_i) + \varepsilon_i, \quad \varepsilon_i | k_i \sim N(0; \sigma_{k_i}^2). \quad (12)$$

For $k_i = 1$, the variance σ_1^2 corresponds to a usual observation, whereas $\sigma_2^2 \gg \sigma_1^2$ corresponds to an outlier. Obviously, (12) is equivalent to the non-Gaussian model

$$y_i = f(x_i) + \varepsilon_i, \quad \varepsilon_i \sim \pi_1 N(0; \sigma_1^2) + \pi_2 N(0; \sigma_2^2), \quad (13)$$

where errors ε_i come from a mixture of normals with $\pi_1 = \text{pr}(k_i = 1)$ and $\pi_2 = 1 - \pi_1 = \text{pr}(k_i = 2)$. The simple model (12) can be generalized

and modified in various ways. For example, (13) can be extended to a finite or continuous mixture of normals. Taking χ^2 -distributed variables as mixture variables leads to observation models with heavy-tailed Student errors, thus making the observation model robust against additive outliers, see, for example, Carter and Kohn (1994), Fahrmeir and Künstler (1998). In principle, conditionally Gaussian observation models can be combined with any of the smoothness priors of Section 2, leading to robust nonparametric regression. Gibbs sampling is ideally suited for Bayesian inference in such models, with an efficient multi-move sampler for Gaussian models as a basic building block. Carter and Kohn (1996) generalize the finite mixture model to the case, where the sequence (k_1, \dots, k_n) is a discrete Markov chain with known transition probabilities. In analogy to the locally dynamic models in Subsection 2.2, Shepard (1994) proposes local scale models with error variables $\varepsilon_i \sim N(0; \sigma_i^2)$, where the variances, perhaps after an appropriate reparameterization, follow random walk models. A special version of such models is the stochastic volatility model $y_i = \varepsilon_i \exp(0.5\alpha_i)$, $\alpha_{i+1} = \phi\alpha_i + v_i$, where y_i are log-returns of a financial assets, and α_i and v_i are mutually independent Gaussian random variables. This non-Gaussian model has been used to generalize the Black-Scholes option pricing equation to allow the volatility change over time.

3.2 Fundamentally non-Gaussian responses

Although a more general formulation is possible, we restrict discussion to nonparametric extensions of generalized linear models. This class of models covers regression situations with responses that are binary, categorical, counts or nonnegative variables. The observation model (1) is replaced by the distributional assumption that, conditional upon covariates and unknowns, the distribution of the response Y belongs to the exponential family. The structural assumption $E(Y_i) = \mu_i = f(x_i)$ in (1) is generalized to

$$\mu_i = h(f(x_i)), \quad i = 1, \dots, n \quad (14)$$

where h is a known response (or link) function. A well known discrete response model is the logit model, where Y is binary or binomial and h is the logistic function $h() = \exp()/[1 + \exp()]$. For the case of several regressors, generalized additive and semiparametric additive models are defined by extending (14) to

$$\mu_i = h\left(\alpha + \sum_{j=1}^p f(x_{ij}) + w_i'\beta\right). \quad (15)$$

For semiparametric Bayesian inference, the two approaches in Section 2, smoothness priors and basis functions, are again suitable, in principle. However, compared to Gaussian and conditionally Gaussian models, much less has been done for regression with fundamentally non-Gaussian responses. A

main feature of Bayesian inference in this situation is that conditional distributions do not have a simple form in general. In particular the conditional distributions for unknown functions f_j are no longer multivariate normal. Hence Gibbs sampling is no longer feasible, and more general Metropolis-Hastings algorithms are needed.

Hastie and Tibshirani (1998) make a corresponding suggestion for extending their smoothness priors approach and “Bayesian backfitting” algorithm to generalized additive models. Fahrmeir and Lang (1998) supplement the observation models (14) or (15) with the Gaussian random walk smoothness priors of Subsection 2.1. They base posterior sampling on an efficient Metropolis-Hastings algorithm recently developed by Knorr-Held (1998) in the context of dynamic generalized linear models. Compared to single move samplers described in Fahrmeir and Tutz (1997, ch. 8), this block move sampler considerably improves convergence and mixing behaviour of posterior samples. Based on the state space representation of stochastic differential priors for smoothing splines, Biller and Fahrmeir (1997) develop Bayesian spline-type smoothing. However, mixing behaviour for generalized additive models is less satisfactory. Surprisingly none of the basis function approaches described in Subsection 2.3 seems to have been extended to fundamentally non-Gaussian responses. The only development at the moment is found in Biller (1998). In an approach related to, yet different from Denison et al. (1998), he works with a B-spline basis, with location and number of knots unknown and estimated via a reversible jump M-H-algorithm, simultaneously with the coefficients of the basis functions.

4 Conclusions

Due to their modular structure, hierarchical Bayesian regression models provide a flexible framework for extensions to other settings. For longitudinal and spatial data, correlation can be accounted for by introducing additional exchangeable random effects or effects with Markov random field priors into the predictor. In particular combining fundamentally non-Gaussian observation models with locally adaptive priors for unsmooth functions as in 2.2, or with basis function priors as in 2.3, an incorporation of unsmooth dynamic and spatial effects offers many possibilities for modelling and analysing data with complex correlation structure in space and time.

References

- BERZUINI, C., LARIZZA, CH. (1996): A Unified Approach for Modeling Longitudinal and Failure Data, With Application in Medical Monitoring. *IEEE Trans. Pattern Analysis and Machine Int.* **18**, 109-123.
- BILLER, C. (1998): Adaptive Bayesian Regression Splines in Semiparametric Generalized Linear Models. *Universität München, Institut für Statistik, SFB 386 Discussion paper 133*.

- BILLER, C., FAHRMEIR, L. (1997): Bayesian Spline-type smoothing in generalized regression models. *Computational Statistics*, **12**, 135-151.
- BOX, G., TIAO, G. (1973): Bayesian Inference in Statistical Analysis. Addison-Wesley, reprinted by Wiley, N.Y., 1992.
- CARLIN, B., POLSON, N. and STOFFER, D. (1992): A Monte Carlo approach to nonnormal and nonlinear state space modelling. *Journal of the American Statist. Ass.* **87**, 493-500.
- CARTER, C., KOHN, R. (1994): On Gibbs sampling for state space models. *Biometrika*, **81**, 541-553.
- CARTER, C., KOHN, R. (1996): Markov chain Monte Carlo in conditionally Gaussian state space models. *Biometrika*, **83**, 3, 589-601.
- DENISON, D., MALLICK, B. and SMITH, A. (1998): Automatic Bayesian curve fitting. *Journal of the Royal Stat. Society*, **B 60**, 333-350.
- DONOHO, D., JOHNSTONE, I. (1994): Ideal spatial adaption by wavelet shrinkage. *Biometrika*, **81**, 425-455.
- FAHRMEIR, L., KAUFMANN, H. (1991): On Kalman filtering, posterior mode estimation and Fisher scoring in dynamic exponential family regression. *Metrika* **38**, 37-60.
- FAHRMEIR, L., TUTZ, G. (1997, 2nd reprint): Multivariate Statistical Modelling based on Generalized Linear Models. Springer, New York.
- FAHRMEIR, L., KÜNSTLER, R. (1998): Penalized Likelihood Smoothing in Robust State Space Models. To appear in *Metrika*.
- FAHRMEIR, L., 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., FAHRMEIR, L. (1998): Function estimation with locally adaptive dynamic models. *Universität München, Institut für Statistik, SFB 386 Discussion paper 135*.
- FRUEWIRTH-SCHNATTER, S. (1994): Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, **15**, 2, 183-202.
- GREEN, P. (1995): Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**, 711-732.
- HASTIE, T., TIBSHIRANI, R. (1998): Bayesian Backfitting. Preprint, Department of Statistics, Stanford.
- HASTIE, T., TIBSHIRANI, R. (1990): Generalized Additive Models. Chapman and Hall
- DEJONG, P., SHEPARD, N. (1995): The simulation smoother for time series models. *Biometrika*, **82**, 339-350.
- KNORR-HELD, L. (1997): Hierarchical Modelling of Discrete Longitudinal Data. Herbert Utz Verlag.

- KNORR-HELD, L. (1998): Conditional Prior Proposals in Dynamic Models. *Scandinavian Journal of Statistics*, to appear.
- SHEPARD, N. (1994): Local scale models. *Journal of Econometrics* **60**, 181-202.
- SMITH, M., KOHN, R. (1996): Nonparametric Regression using Bayesian variable selection. *Journal of Econometrics* **75**, 317-343.
- WHITTAKER, E.T. (1923): On a New Method of Graduation. *Proc. Edinburgh Math. Ass.* **78**, 81-89.
- WONG, C., KOHN, R. (1996): A Bayesian approach to estimating and forecasting additive nonparametric autoregressive models. *Journal of Time Series Analysis*, **17**, 203-220.