Smoothing Spline Nonlinear Nonparametric Regression Models

Yuedong Wang
Department of Statistics and Applied Probability
University of California-Santa Barbara

Chunlei Ke
Amgen

Outline

Outline

Outline

Outline

Outline

Smoothing spline nonparametric regression

A standard smoothing spline model assumes that

$$y_i = f(t_i) + \epsilon_i, \quad i = 1, \cdots, n$$

- y_i are observations
- f is an unknown function belonging to a model space
- t_i are design points
- ϵ_i are random errors with $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

General smoothing spline nonparametric regression

In many applications observations of the mean function are made indirectly. Wahba (1990) considered the following general smoothing spline model

$$y_i = \mathcal{L}_i f + \epsilon_i, \quad i = 1, \cdots, n$$

- y_i are observations
- f is observed through a known bounded linear functional $\mathcal{L}_i$
- Simple smoothing spline model is a special case with $\mathcal{L}_i f = f(t_i)$, $\mathcal{L}_i$ is called an evaluation functional
- Other interesting examples of $\mathcal{L}_i$ are $\mathcal{L}_i f = \int_a^b w_i(t) f(t) dt$ and $\mathcal{L}_i f = f'(t_i)$
- ϵ_i are random errors with $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

Motivations

The nonlinear nonparametric regression is necessary because

- in some experiments f may only be observed indirectly through a nonlinear functional;
- nonlinear transformations are useful tools to relax constraints on f.

Motivations

The nonlinear nonparametric regression is necessary because

- in some experiments f may only be observed indirectly through a nonlinear functional;
- nonlinear transformations are useful tools to relax constraints on f.

Example 1: Remote sensing

The satellite up-welling radiance measurements R_v are related to the underlying atmospheric temperature distribution g through

$$R_v(g) = B_v(g(x_s))\tau_v(x_s) - \int_{x_0}^{x_s} B_v(g(x))\tau_v'(x)dx$$

- x is some monotone transformation of pressure p
- $\tau_v(x)$ is the transmittance of the atmosphere above x at wavenumber v
- $B_v(g) = c_1 v^3/[exp(c_2 v/g) - 1]$ with known constants c_1 and c_2
- The goal is to estimate g as a function of x using noisy observations of $R_v(g)$
- $R_v(g)$ is nonlinear in g

Example 2: Positive inverse problem

For simplicity, suppose that observations are generated by the Fredholm's integral equation of the first kind

$$y_i = \int K(t_i, s) f(s) ds + \epsilon_i, \quad i = 1, \cdots, n$$

- K is a known impulse response function
- ϵ_i are measurement errors
- The goal is to recover f through observations
- Often f is known to be positive. We can relax the positive constraint by a simple transformation $f = \exp(g)$ and then estimate g. The transformed model is nonlinear in g

Example 3: Extensions of additive models

An additive model assumes that

$$y_i = \alpha + f_1(t_{1i}) + \cdots + f_r(t_{ri}) + \epsilon_i, \quad i = 1, \cdots, n$$

A simple extension is to allow nonlinear functionals for some or all components:

$$y_i = \alpha + \mathcal{N}_i^1 g_1 + \cdots + \mathcal{N}_i^r g_r + \epsilon_i, \quad i = 1, \cdots, n,$$

where $\mathcal{N}^k$ are known linear or nonlinear operators.

For example, if f_1 is known to be strictly increasing, then $f_1'(t) > 0$. Let $f_1'(t) = \exp(g_1(t))$. We can re-express f_1 as $f_1(t) = f_1(0) + \int_0^t \exp(g_1(s))ds$. The constant $f_1(0)$ is absorbed by α. Therefore we have $\mathcal{N}^1 g_1 = \int_0^{x_1} \exp(g_1(t))dt$.

Example 3: Extensions of additive models

An additive model assumes that

$$y_i = \alpha + f_1(t_{1i}) + \cdots + f_r(t_{ri}) + \epsilon_i, \quad i = 1, \cdots, n$$

A simple extension is to allow nonlinear functionals for some or all components:

$$y_i = \alpha + \mathcal{N}_i^1 g_1 + \cdots + \mathcal{N}_i^r g_r + \epsilon_i, \quad i = 1, \cdots, n,$$

where $\mathcal{N}^k$ are known linear or nonlinear operators.

For example, if f_1 is known to be strictly increasing, then $f_1'(t) > 0$. Let $f_1'(t) = \exp(g_1(t))$. We can re-express f_1 as $f_1(t) = f_1(0) + \int_0^t \exp(g_1(s))ds$. The constant $f_1(0)$ is absorbed by α. Therefore we have $\mathcal{N}^1 g_1 = \int_0^{x_1} \exp(g_1(t))dt$.

Nonlinear nonparametric regression models

We define a general smoothing spline nonlinear nonparametric regression model (SSNNRM) as

$$y_i = \mathcal{N}_i(g_1, \cdots, g_r) + \epsilon_i, \quad i = 1, \cdots, n$$

- y_i are observations
- $\mathcal{N}_i$ are known nonlinear functionals
- $g_1, \cdots, g_r$ are unknown functions
- ϵ_i are random errors with $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

Model spaces

To deal with different kind of covariates in a unified fashion, we assume that, for each $k = 1, \cdots, r$,

- The domain $\mathcal{T}_k$ for function g_k is an arbitrary set
- $g_k \in \mathcal{H}_k$, where $\mathcal{H}_k$ is a *reproducing kernel Hilbert space* (RKHS) on a domain $\mathcal{T}_k$
- Examples of model spaces and domains are *polynomial splines* on $[0, 1]$, *periodic spline* on the unit circle, *thin-plate splines* on Euclidean d-space, and *tensor product splines* on tensor products of domains

Constructions of model spaces

We further assume that

$$\mathcal{H}_k = \mathcal{H}_{k0} \oplus \mathcal{H}_{k1}$$

where

$$\mathcal{H}_{k0} = span\{\phi_{k1}(t), \cdots, \phi_{km_k}(t)\}$$

contains functions which are not penalized, and $\mathcal{H}_{k1}$ is a RKHS with reproducing kernel (RK) $R_{k1}(s,t)$.
Denote P_{k1} as the projection operator onto the subspace $\mathcal{H}_{k1}$ in $\mathcal{H}_k$. Then $||P_{k1}g_k||^2$ measures the departure of g_k from $\mathcal{H}_{k0}$.
The choices of $\mathcal{H}_{k0}$, $\mathcal{H}_{k1}$ and P_{k1} depend on

- domain of the function $\mathcal{T}$
- prior knowledge such as smoothness
- purpose of the study

For example, for a *cubic spline*, $\mathcal{H}_{k0} = span\{1, t\}$,
$R_{k1}(s,t) = \int_0^1 (s-u)_+(t-u)_+ du$, $||P_{k1}g_k||^2 = \int_0^1 (g_k''(s))^2 ds$.

Penalized least squares

We estimate $g_1, \cdots, g_r$ by minimizing the following penalized least squares

$$PLS = \sum_{i=1}^{n} (y_i - \mathcal{N}_i(g_1, \cdots, g_r))^2 + n\lambda \sum_{k=1}^{r} \theta_k ||P_{k1} g_k||^2$$

- The first part measures the goodness-of-fit
- The second part is a penalty to the departure from spaces $\mathcal{H}_{k0}$
- λ and θ_k are smoothing parameters which balance the trade-off between the goodness-of-fit and the penalty
- Under some regularity conditions, the PLS has a unique minimizer
- For linear operators, the solutions to the PLS fall into a finite dimensional space. No longer holds in general for SSNNRMs

Exact solution for a special case

Assume that $\mathcal{N}_i(g_1, \cdots, g_r)$ depends on g_k through $\mathcal{L}_{ki} g_k$ only

$$\mathcal{N}_i(g_1, \cdots, g_r) = \eta_i(\mathcal{L}_{1i} g_1, \cdots, \mathcal{L}_{ri} g_r), \quad i = 1, \cdots, n$$

- η_i are known nonlinear functions
- $\mathcal{L}_{1i}, \cdots, \mathcal{L}_{ri}$ are bounded linear operators

Theorem. The solution to the PLS can be represented as

$$\hat{g}_k(t) = \sum_{i=1}^{m_k} d_{ki} \phi_{ki}(t) + \sum_{j=1}^{n} \theta_k c_{kj} \xi_{kj}(t), \quad k = 1, \cdots, r$$

- $\xi_{ki} = P_{k1} \mathcal{L}_{ki} R_{k1}$
- We need to solve coefficients d_{ki} and c_{kj}. For fixed smoothing parameters, standard nonlinear optimization procedures such as the Gauss-Newton and Newton-Raphson methods can be used

Approximation

- For the general SSNNRM the solutions to the PLS do not fall into a finite dimensional space
- We first consider the case of $r = 1$
- Let g_- be current estimate of g. Assume that the Fréchet differential of $\mathcal{N}_i$ with respect to g evaluated at g_- exists and is bounded
- Denote

$$\mathcal{D}_i = \partial \mathcal{N}_i / \partial g|_{g=g_-}$$

- We approximate $\mathcal{N}_i g$ by its first order Taylor expansion at g_-:

$$\mathcal{N}_i g \approx \mathcal{N}_i g_- + \mathcal{D}_i (g - g_-)$$

Extended Gauss-Newton algorithm

Then we can approximate the original SSNNRM by

$$\tilde{y}_i = \mathcal{D}_i g + \epsilon_i, \quad i = 1, \cdots, n$$

where $\tilde{y}_i = y_i - \mathcal{N}_i g_- + \mathcal{D}_i g_-$.

We minimize

$$\sum_{i=1}^{n} (\tilde{y}_i - \mathcal{D}_i g)^2 + n\lambda ||P_1 g||^2$$

to get a new estimate of g.

- Since $\mathcal{D}_i$ is a linear and bounded functional, the solution to the above PLS falls in a finite dimensional space
- Thus the solution can be represented as a linear combination of basis and representers. Coefficients can be solved by a nonlinear optimization procedure

Backfitting

When $r > 1$, we use a Gauss-Seidel-type algorithm to estimate functions iteratively one at a time.

Algorithm

1. Initialize: $g_i = g_i^0, \quad i = 1, \cdots, r$
2. Cycle: for $k = 1, \cdots, r, 1, \cdots, r, \cdots$, conditional on the current estimates of $g_1, \cdots, g_{k-1}, g_{k+1}, \cdots, g_r$, update g_k as the minimizer of

$$\sum_{i=1}^{n} (y_i - \mathcal{N}_i(g_1, \cdots, g_{k-1}, g_k, g_{k+1}, \cdots, g_r))^2 + n\lambda ||P_{k1} g_k||^2$$

3. Continue step (2) until convergence

Smoothing parameters and inference

- As usual, the smoothing parameters are critical to the performance of spline estimates
- The generalized cross validation (GCV) and the generalized maximum likelihood (GML) methods were extended to estimate smoothing parameters in SSNNRMs
- Smoothing parameters are estimated iteratively at each iteration
- Bayes models were constructed for SSNNRMs which allows us to construct approximate Bayesian confidence intervals
- Bootstrap method can also be used to construct confidence intervals
- The estimation and inference methods performed well in simulations

Smoothing parameters and inference

- As usual, the smoothing parameters are critical to the performance of spline estimates
- The generalized cross validation (GCV) and the generalized maximum likelihood (GML) methods were extended to estimate smoothing parameters in SSNNRMs
- Smoothing parameters are estimated iteratively at each iteration
- Bayes models were constructed for SSNNRMs which allows us to construct approximate Bayesian confidence intervals
- Bootstrap method can also be used to construct confidence intervals
- The estimation and inference methods performed well in simulations

Smoothing parameters and inference

- As usual, the smoothing parameters are critical to the performance of spline estimates
- The generalized cross validation (GCV) and the generalized maximum likelihood (GML) methods were extended to estimate smoothing parameters in SSNNRMs
- Smoothing parameters are estimated iteratively at each iteration
- Bayes models were constructed for SSNNRMs which allows us to construct approximate Bayesian confidence intervals
- Bootstrap method can also be used to construct confidence intervals
- The estimation and inference methods performed well in simulations

Smoothing parameters and inference

- As usual, the smoothing parameters are critical to the performance of spline estimates
- The generalized cross validation (GCV) and the generalized maximum likelihood (GML) methods were extended to estimate smoothing parameters in SSNNRMs
- Smoothing parameters are estimated iteratively at each iteration
- Bayes models were constructed for SSNNRMs which allows us to construct approximate Bayesian confidence intervals
- Bootstrap method can also be used to construct confidence intervals
- The estimation and inference methods performed well in simulations

Smoothing parameters and inference

- As usual, the smoothing parameters are critical to the performance of spline estimates
- The generalized cross validation (GCV) and the generalized maximum likelihood (GML) methods were extended to estimate smoothing parameters in SSNNRMs
- Smoothing parameters are estimated iteratively at each iteration
- Bayes models were constructed for SSNNRMs which allows us to construct approximate Bayesian confidence intervals
- Bootstrap method can also be used to construct confidence intervals
- The estimation and inference methods performed well in simulations

Smoothing parameters and inference

- As usual, the smoothing parameters are critical to the performance of spline estimates
- The generalized cross validation (GCV) and the generalized maximum likelihood (GML) methods were extended to estimate smoothing parameters in SSNNRMs
- Smoothing parameters are estimated iteratively at each iteration
- Bayes models were constructed for SSNNRMs which allows us to construct approximate Bayesian confidence intervals
- Bootstrap method can also be used to construct confidence intervals
- The estimation and inference methods performed well in simulations

Cross validation

Let $g_1^{[v]}, \cdots, g_r^{[v]}$ be minimizers of of the PLS without the vth observation

$$\sum_{i \neq v} (y_i - \mathcal{N}_i(g_1, \cdots, g_r))^2 + n\lambda \sum_{k=1}^{r} \theta_k ||P_{k1} g_k||^2$$

The ordinary leave-out-one cross validation method selects $\lambda/\theta_1, \cdots, \lambda/\theta_r$ as minimizers of the following score

$$OCV(\lambda, \boldsymbol{\theta}) = \frac{1}{n} \sum_{v=1}^{n} (y_v - \mathcal{N}_v(g_1^{[v]}, \cdots, g_r^{[v]}))^2$$

Leaving-out-one lemma

Lemma For fixed v and z, let $\boldsymbol{h}[v,z]$ be the vector of functions that minimizes

$$(z-\mathcal{N}_v(g_1,\cdots,g_r))^2+\sum_{i\neq v}(y_i-\mathcal{N}_i(g_1,\cdots,g_r))^2+n\lambda\sum_{k=1}^{r}\theta_k||P_{1k}g_k||^2$$

Then

$$\boldsymbol{h}[v,\mathcal{N}_v(g_1^{[v]},\cdots,g_r^{[v]})]=(g_1^{[v]},\cdots,g_r^{[v]})$$

Generalized cross validation

The OCV criterion can be approximated by

$$V(\lambda, \boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mathcal{N}_i(g_1, \cdots, g_r))^2 / \left[\frac{1}{n} tr(\boldsymbol{I} - \boldsymbol{A})\right]^2$$

- $a_{ij} = \partial \mathcal{N}_i(\boldsymbol{h}[i, y_j])/\partial y_j = \sum_{u=1}^{r} \frac{\partial \mathcal{N}_i(g_1, \cdots, g_r)}{\partial g_u} \frac{\partial g_u}{\partial y_j}$
- $\boldsymbol{A} = (a_{ij})_{i,j=1}^{n}$
- Function V is called the GCV criterion. Its minimizer is called the GCV estimate of the smoothing parameter
- It is difficult to compute a_{ij} directly due to nonlinear functionals
- We approximate a_{ij} by replacing g_i with their current estimates
- This leads to an iterative procedure which estimates smoothing parameters at each iteration

Linear approximation

At each iteration, the SSNNRM is approximated by

$$\tilde{y}_i = \mathcal{D}_i g + \epsilon_i, \quad i = 1, \cdots, n$$

- $\mathcal{D}_i = \partial \mathcal{N}_i / \partial g|_{g=g_-}$
- $\tilde{y}_i = y_i - \mathcal{N}_i g_- + \mathcal{D}_i g_-$

Bayes model

Assume a prior distribution for g as

$$G(t) = \sum_{i=1}^{m} d_i \phi_i(t) + \tau^{1/2} Z(t)$$

- $d_i \overset{iid}{\sim} \mathcal{N}(0, a)$
- $Z(t)$ is a mean zero Gaussian process with $\text{Cov}(Z(s), Z(t)) = R_1(s, t)$

Assume that observations are generated from the following model

$$\tilde{y}_i = \mathcal{D}_i G + \epsilon_i, \quad i = 1, \cdots, n$$

Bayes model

Assume a prior distribution for g as

$$G(t) = \sum_{i=1}^{m} d_i \phi_i(t) + \tau^{1/2} Z(t)$$

- $d_i \stackrel{iid}{\sim} \mathcal{N}(0, a)$
- $Z(t)$ is a mean zero Gaussian process with $\text{Cov}(Z(s), Z(t)) = R_1(s, t)$

Assume that observations are generated from the following model

$$\tilde{y}_i = \mathcal{D}_i G + \epsilon_i, \ \ i = 1, \cdots, n$$

Generalize Maximum Likelihood (GML) method

- Since $\mathcal{D}_i$ are linear operators, the posterior mean of the Bayes model equals the solution to the PLS at this iteration
- Similar arguments as in the linear case leads to the following GML criterion

$$M(\lambda, \boldsymbol{\theta}) = \frac{\frac{1}{n}\sum_{i=1}^{n}(y_i - \mathcal{N}_i(g_1, \cdots, g_r))^2}{\left[\det^+(I - \boldsymbol{A})\right]^{1/(n-\sum_{k=1}^{r} m_k)}}$$

  where $\det^+$ is the product of nonzero eigenvalues
- The minimizer to the GML criterion is called the GML estimates of the smoothing parameters
- Since a_{ij} are calculated based on current estimates, as GCV, the GML is an iterative procedure

Linear approximation at convergence

At convergence, the SSNNRM is approximated by

$$\tilde{y}_i^* = \mathcal{D}_i^* g + \epsilon_i, \quad i = 1, \cdots, n$$

- $\hat{g}$ is the estimate at convergence
- $\mathcal{D}_i^* = \partial \mathcal{N}_i / \partial g|_{g=\hat{g}}$
- $\tilde{y}_i^* = y_i - \mathcal{N}_i \hat{g} + \mathcal{D}_i^* \hat{g}$

Bayesian model at convergence

Assume a prior distribution for g as

$$G(t) = \sum_{i=1}^{m} d_i \phi_i(t) + \tau^{1/2} Z(t)$$

- $d_i \overset{iid}{\sim} \mathcal{N}(0, a)$
- $Z(t)$ is a mean zero Gaussian process with $\text{Cov}(Z(s), Z(t)) = R_1(s, t)$

Assume that observations are generated from the following model

$$\tilde{y}_i^* = \mathcal{D}_i^* G + \epsilon_i, \quad i = 1, \cdots, n$$

Bayesian model at convergence

Assume a prior distribution for g as

$$G(t) = \sum_{i=1}^{m} d_i \phi_i(t) + \tau^{1/2} Z(t)$$

- $d_i \overset{iid}{\sim} \mathcal{N}(0, a)$
- $Z(t)$ is a mean zero Gaussian process with $\text{Cov}(Z(s), Z(t)) = R_1(s, t)$

Assume that observations are generated from the following model

$$\tilde{y}_i^* = \mathcal{D}_i^* G + \epsilon_i, \quad i = 1, \cdots, n$$

Bayesian confidence intervals

- Since $\mathcal{D}_i^*$'s are linear operators, the posterior mean of the Bayesian model equals $\hat{g}(t)$
- Posterior variances and Bayesian confidence intervals can be calculated
- Bayesian confidence intervals can constructed using posterior variances
- The performance of Bayesian confidence intervals depends largely on the accuracy of the linear approximation. When curvature of $\mathcal{N}_i$ with respect to g is high, modification is necessary to improve coverage
- Simulations indicate that Bayesian confidence intervals perform reasonably well when curvature is not high
- Bootstrap confidence intervals can also be constructed

R package

We have developed a generic R function, `nnr`, for fitting SSNNRMs. `nnr` is one function in the ASSIST package which can be downloaded from

http://cran.r-project.org/

Details and examples can be found in the manual of the ASSIST package downloadable from

http://www.pstat.ucsb.edu/faculty/yuedong

Fit a positive function

Consider a nonparametric regression model

$$y_i = f(t_i) + \epsilon_i, \quad t_i \in [0,1], \quad i = 1, \cdots, n$$

- Positive constraint: $f > 0$
- We may use one of the following two transformations to enforce positivity:
  - exponential transformation $f = \exp(g)$
  - square transformation $f = g^2$
- Then we can model the unconstrained function g by a spline model

A simple simulation

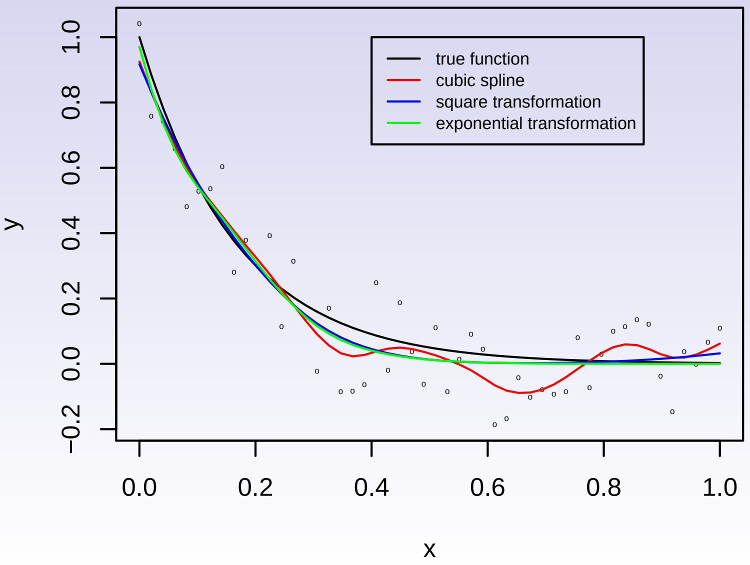
```
n <- 50
x <- seq(0,1,len=n)
y <- exp(-6*x)+.1*rnorm(n)

# fit a cubic spline
ssrfit <- ssr(y~x, cubic(x))

# fit the square transformed model
nnrfit1 <- nnr(y~g(x)**2,
               func=g(u)~list(~u,cubic(u)),
               start=list(g=sqrt(abs(y))))

# fit the exponential transformed model
nnrfit2 <- nnr(y~exp(g(x)),
               func=g(u)~list(~u,cubic(u)),
               start=list(g=log(abs(y)+0.001)))
```

Fit


true function
cubic spline
square transformation
exponential transformation
y
x
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0

Term structure of interest rates

The term structure of interest rates is a concept central to economic and financial theory. Consider a set of n coupon bonds from which the interest rate term structure is to be inferred. Denote the current time as zero. Then the pricing model is

$$y_i = \sum_{j=1}^{m_i} c_{ij}\delta(t_{ij}) + \epsilon_i, \quad i = 1, \cdots, n$$

- y_i is the current price of bond i
- c_{ij} is the payment paid at a future time t_{ij}, $0 < t_{i1} < \cdots < t_{im_i}$
- $\delta(t)$ is the discount function: price of a dollar delivered at date t
- ϵ_i are iid random errors with mean zero and variance σ^2

Constraints

The discount function δ is required to be

- $\delta(0) = 1$
- positive
- decreasing

It is difficult to estimate δ directly due to these constraints. Often δ is represented by

$$\delta(t) = \exp(-\int_0^t g(s)ds)$$

- The transformation takes care of all constraints on δ
- $g(s) \geq 0$ is called the forward rate
- Replacing δ by its transformation leads to a SSNNRM with $\mathcal{N}_i g = \sum_{j=1}^{m_i} c_{ij} \exp(-\int_0^{t_{ij}} g(s)ds)$

Treasury and corporate bonds

- While the estimation of Treasury term structure has received enormous attention, little research has been done for corporate bonds.
- Special estimation procedures is necessary due to the lack of data for corporate bond prices.
- To borrow information from Treasury bonds, Jarrow, Ruppert and Yu (2004) modelled the forward rate of corporate bonds as a Treasure bond plus a parametric credit spread. A polynomial function of low degree was used to model the credit spread.
- We now illustrate how to use SSNNRM to investigate differences between two groups of bonds. We model the credit spread nonparametrically. Thus our methods may be used to check a parametric model.

Treasury and corporate bonds

- While the estimation of Treasury term structure has received enormous attention, little research has been done for corporate bonds.
- Special estimation procedures is necessary due to the lack of data for corporate bond prices.
- To borrow information from Treasury bonds, Jarrow, Ruppert and Yu (2004) modelled the forward rate of corporate bonds as a Treasure bond plus a parametric credit spread. A polynomial function of low degree was used to model the credit spread.
- We now illustrate how to use SSNNRM to investigate differences between two groups of bonds. We model the credit spread nonparametrically. Thus our methods may be used to check a parametric model.

Treasury and corporate bonds

- While the estimation of Treasury term structure has received enormous attention, little research has been done for corporate bonds.
- Special estimation procedures is necessary due to the lack of data for corporate bond prices.
- To borrow information from Treasury bonds, Jarrow, Ruppert and Yu (2004) modelled the forward rate of corporate bonds as a Treasure bond plus a parametric credit spread. A polynomial function of low degree was used to model the credit spread.
- We now illustrate how to use SSNNRM to investigate differences between two groups of bonds. We model the credit spread nonparametrically. Thus our methods may be used to check a parametric model.

Treasury and corporate bonds

- While the estimation of Treasury term structure has received enormous attention, little research has been done for corporate bonds.
- Special estimation procedures is necessary due to the lack of data for corporate bond prices.
- To borrow information from Treasury bonds, Jarrow, Ruppert and Yu (2004) modelled the forward rate of corporate bonds as a Treasure bond plus a parametric credit spread. A polynomial function of low degree was used to model the credit spread.
- We now illustrate how to use SSNNRM to investigate differences between two groups of bonds. We model the credit spread nonparametrically. Thus our methods may be used to check a parametric model.

Smoothing Spline ANOVA decomposition

Let $g(k,t)$ be the forward rate for bonds in group k, $k = 1, 2$. We model the group effect using a one-way ANOVA model and the time effect using a cubic spline model. Then we have the following SS ANOVA decomposition

$$g(k,t) = \mu + \alpha_k + \beta t + s_2(t) + \gamma_k t + s_{12}(k,t)$$

- μ is a constant
- α_k is the main effect of group
- βt is the linear main effect of time
- $s_2(t)$ is the smooth main effect of time
- $\gamma_k t$ is the smooth-linear interaction between group and time
- $s_{12}(k,t)$ is the smooth-smooth interaction between group and time

Credit spread

The credit spread equals

$$g(2,t) - g(1,t) = (\alpha_2 - \alpha_1) + (\gamma_2 - \gamma_1)t + (s_{12}(2,t) - s_{12}(1,t))$$

- A constant spread is equivalent to that both interactions $\gamma_k t$ and $s_{12}(k,t)$ equal zero
- A linear spread is equivalent to that the smooth-smooth interaction $s_{12}(k,t)$ equals zero

Data

144 GE (General Electronic Company) bonds and 78 Treasury bonds were collected from Bloomberg. As expected, the GE discount rate is consistently smaller than that of Treasury bonds, representing a higher risk associated with corporate bonds. The credit spread is not significantly different from a constant function which confirms the results in Jarrow, Ruppert and Yu (2004).

Fit

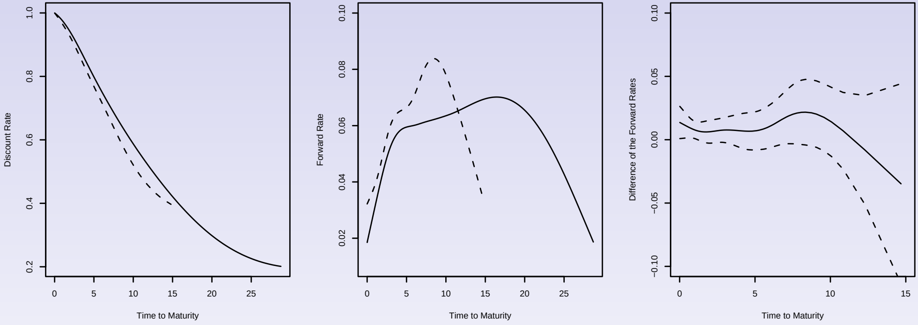
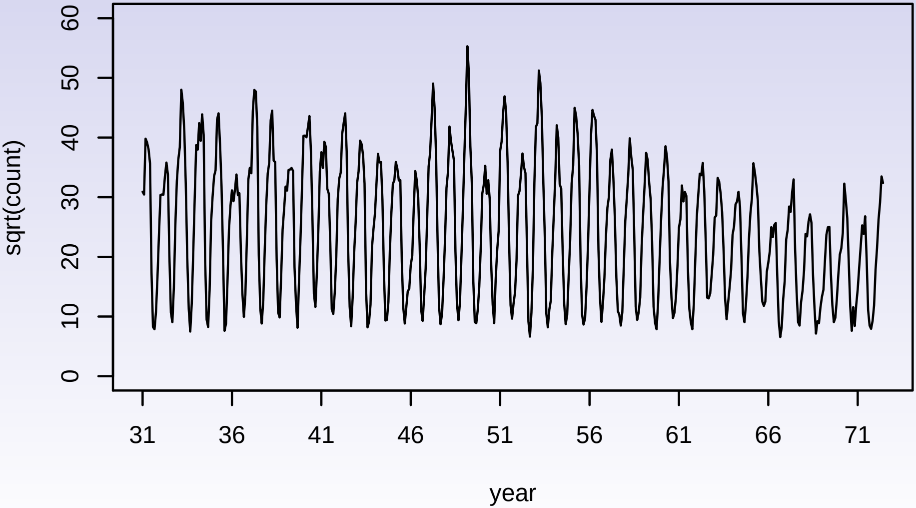


Figure: Estimate of the discount functions (left), forward rates (middle) and difference of forward rates between Treasury and GE bonds (right). On the left and middle panels, solid and dotted lines represent Treasury and GE bonds respectively. On the right panel, solid line represents estimate of the credit spread and dotted lines represent 95% bootstrap confidence intervals.

Chickenpox epidemic

The data set contains monthly number of reported cases of chickenpox in New York City from 1931 to the first six months of 1972. The goal is to investigate dynamics in an epidemic: long term trend over years, seasonal trend and their interactions.

Time series plot



Seasonal variation

- The seasonal variation was mainly caused by two factors:
  - social behavior of children who made close contacts when school was in session
  - temperature and humidity which may affect the survival and transmission of dispersal stages
- Thus the seasonal variations were similar over the years
- We assume that the seasonal variation has the same shape after vertical shift and vertical scale transformations

A multiplicative model for chickenpox epidemic

We assume the following multiplicative model

$$y(t_1, t_2) = g_1(t_2) + \exp(g_2(t_2)) \times g_3(t_1) + \epsilon(t_1, t_2)$$

- $y(t_1, t_2)$ is the square root of reported cases in month t_1 of year t_2
- Both t_1 and t_2 are transformed into the interval $[0, 1]$
- g_1 represents yearly mean cases
- g_2 represents magnitude of the seasonal variation. $\exp(g_2(t_2))$ is the amplitude. A bigger amplitude corresponds to a bigger seasonal variation
- g_3 represents seasonal trend
- All component functions have nice interpretations
- g_1, g_2 and g_3 are unknown and to be modeled nonparametrically
- A extension of the additive models and varying coefficient models

Identifiability and model spaces

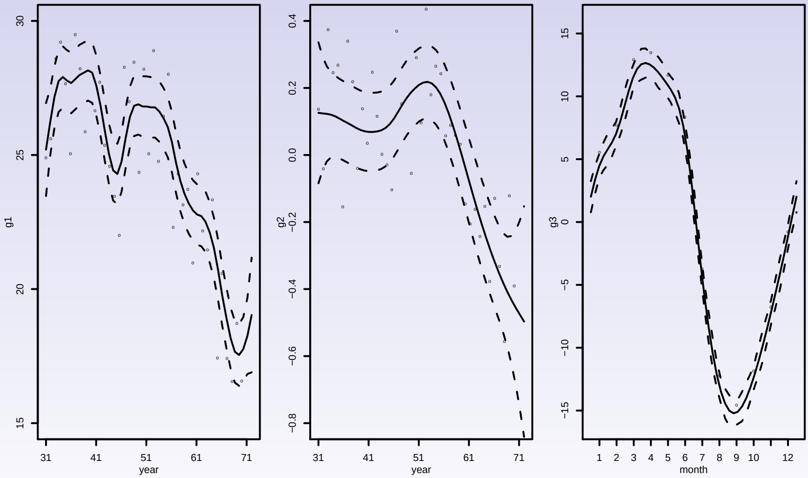
- g_1 is modeled using a cubic spline: $g_1 \in W_2[0,1]$
- The exponential transformation of g_2 makes the amplitude positive. Again, g_2 is modeled using a cubic spline. To make g_2 and g_3 identifiable, we need the side condition $\int_0^1 g_2(t)dt = 0$. We achieve this by removing the constant functions form the model space: $g_2 \in W_2[0,1] \ominus \{1\}$
- It has been recognized g_3 periodic and is close to a sinusoidal function, but a simple sinusoidal model may be inappropriate. We use the L-spline with $L = D^2 + (2\pi)^2$ to model g_3. To make model g_3 identifiable with g_1, we need the side condition $\int_0^1 g_3(t)dt = 0$. Again, we achieve this by removing the constant functions form the model space: $g_3 \in W_2(per) \ominus \{1\}$ where

$$\begin{aligned} W_2(per) &= \{f: \ f \text{ and } f' \text{ are absolutely continuous}, \\ & \qquad f(0)=f(1),\ f'(0)=f'(1),\ \int_0^1 (f'')^2 < \infty\} \end{aligned}$$

R code

```
S3 <- periodic(chickenpox$csmonth)
f3.tmp <- ssr(ct~1,rk=S3,data=chickenpox,
              spar=''m'')
f3.ini <- as.vector(S3%*%f3.tmp$rkpk.obj$c)
nnr(ct~f1(csyear)+exp(f2(csyear))*f3(csmonth),
       func=list(f1(x)~list(~I(x-.5),cubic(x)),
                 f2(x)~list(~I(x-.5)-1,cubic(x)),
                 f3(x)~list(~sin(2*pi*x)+
       cos(2*pi*x)-1,lspline(x,type=''sine0''))),
       data=chickenpox,
       start=list(f1=mean(sqrt(count)),
                  f2=0,f3=f3.ini),
       control=list(converg=''coef''), spar=''m'')
```

Fits



g1
30
25
20
15
31
41
51
61
71
year
g2
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
31
41
51
61
71
year
g3
15
10
5
0
−5
−10
−15
1 2 3 4 5 6 7 8 9 10 12
month

Conclusions

- The SSNNRM is a versatile family of models which may be useful for
    - estimating nonparametric functions when they are observed indirectly throght a nonlinear functional
    - relaxing constratins using nonlinear transformations
    - checking nonlinear regression models
- Gauss-Newton and Gauss-Seidel algorithms were exteded for estimation
- Methods for selecting smoothing parameters and inference were developed
- A user friendly R function, `nnr`, can be used to fit SSNNMs