Penalized Splines and Financial Market Data

David Ruppert
Operations Research and Industrial Engineeering
Cornell University

dr24@cornell.edu
http://legacy.orie.cornell.edu/~davidr

Main Themes

- Calibration of financial models is a statistical problem
- Researchers in mathematical finance are experts in probability theory but often are less knowledgeable about statistical modeling and data analysis
- Unfortunately, statisticians have, with some notable exceptions, not recognized finance as an important area of application
- Transformation and weighting in regression can improve the calibration of financial models
- Splines are an effective tool for data analysis and statistical modeling

Overview

- Recent example where a statistician could have helped
- Example of curve fitting – dynamics of interest rates
- Penalized splines
- Two examples:
  - Return to interest rate dynamics
  - Term structure – estimating the forward rate curve

Example: Estimation of Default Probabilities

Data:

- ratings: 1 = Aaa (best), . . . , 16 = B3 (worse)
- default frequency: estimate of default probability
  - many zero values at best ratings

From recent book on credit risk

- **nonlinear model:**

$$\Pr(\text{default}|\text{rating}) = \exp\{\beta_0 + \beta_1 \text{rating}\}$$

- **linear/transformation model (in recent textbook):**

$$\log\{\Pr(\text{default}|\text{rating})\} = \beta_0 + \beta_1 \text{rating}$$

  – **Problem:** cannot take logs of default frequencies that are 0

  – **(Sub-optimal) solution in textbook:** throw out these observations

- **Transform-both-sides (TBS) model** – see Carroll and Ruppert (1984, 1988):

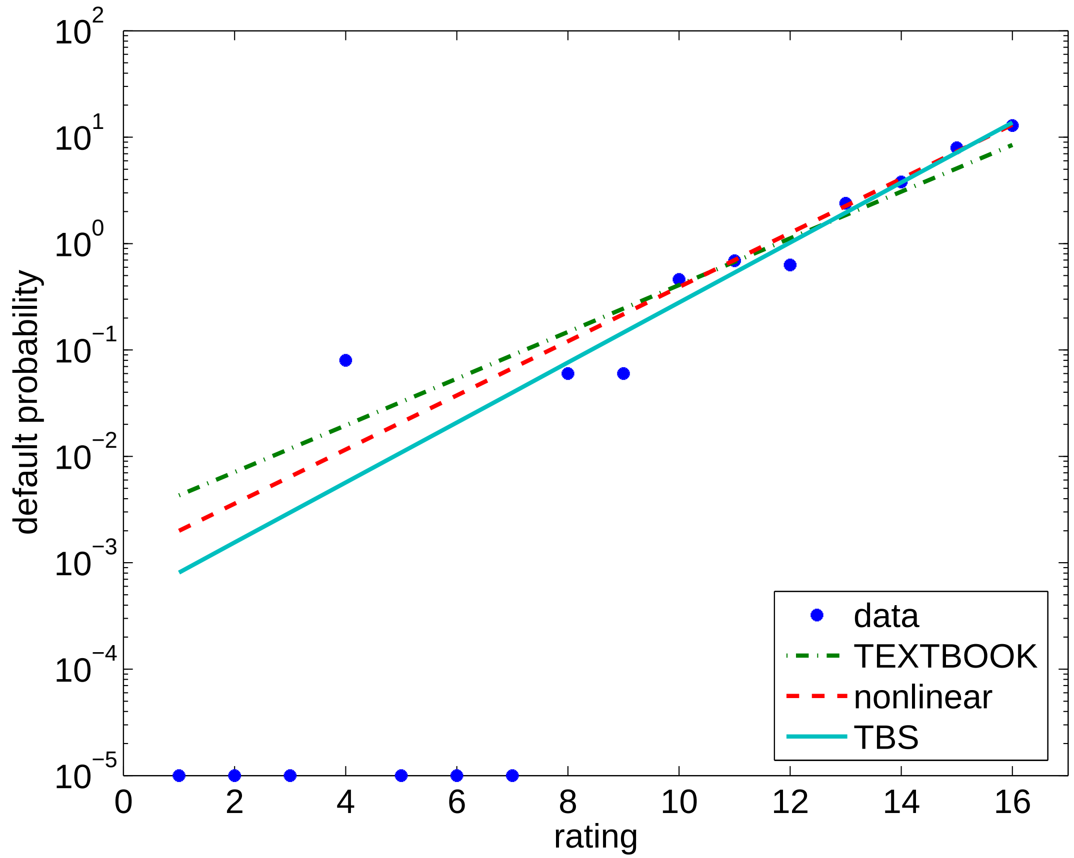
  $$\Pr(\text{default}|\text{rating})^{\alpha} = \exp[\alpha\{\beta_0 + \beta_1 \text{rating}\}]$$

  – α chosen by residual plots (or maximum likelihood)

  – $\alpha = 1/2$ works well

  – $\alpha = 0 \Rightarrow$ log transformation

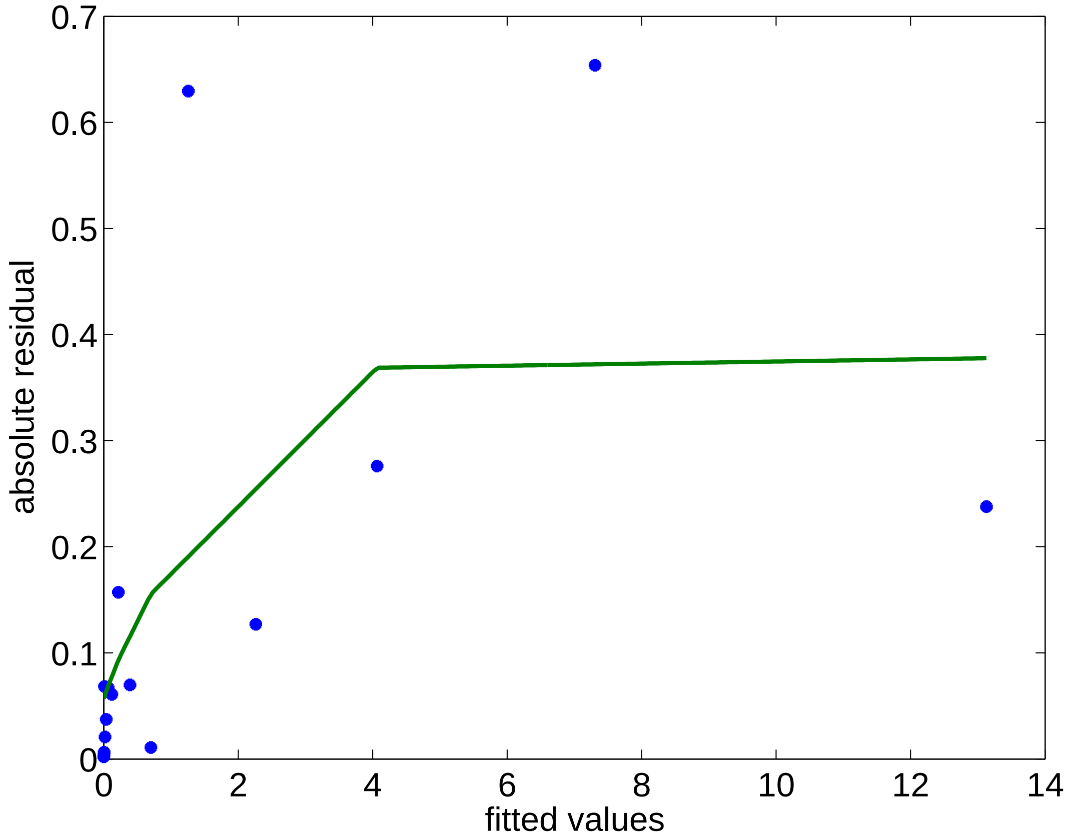
    * if we $x \mapsto x^{\alpha}$ by $x \mapsto (x^{\alpha} - 1)/\alpha$



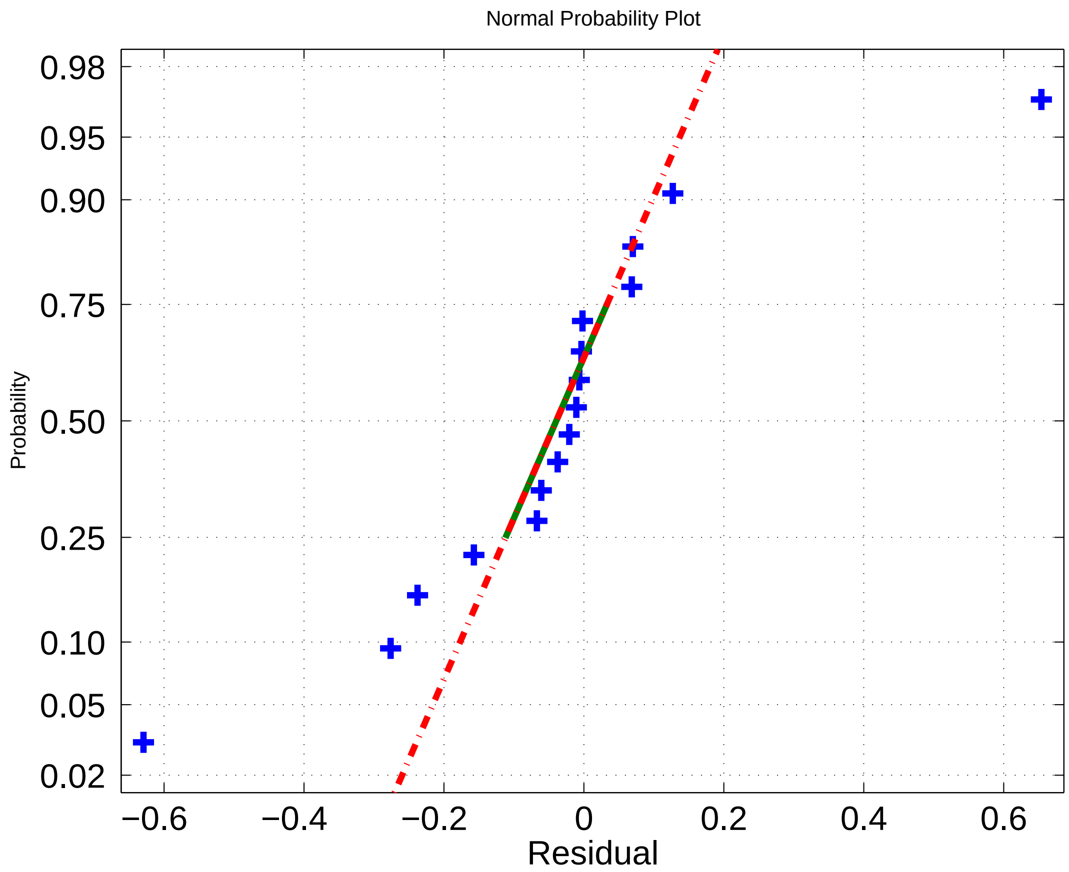
TBS fit compared to others

Data = proportion defaulting

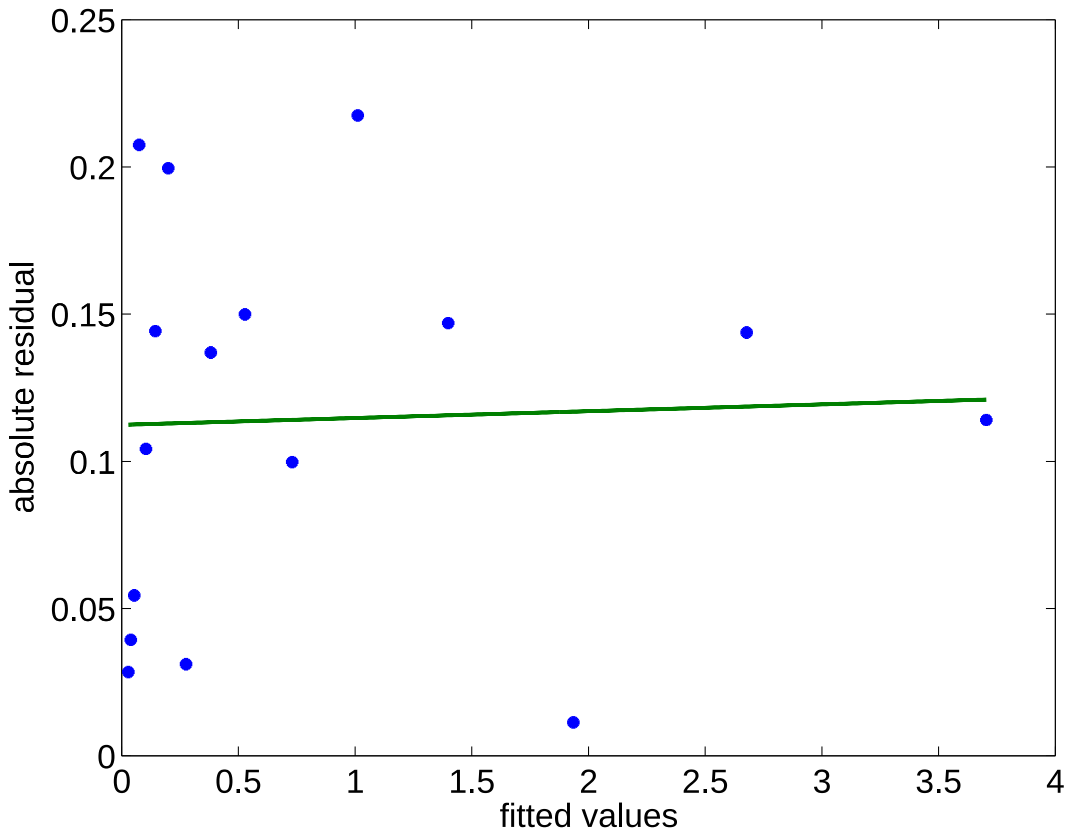
Values at bottom are at log(proportion) $= -\infty$



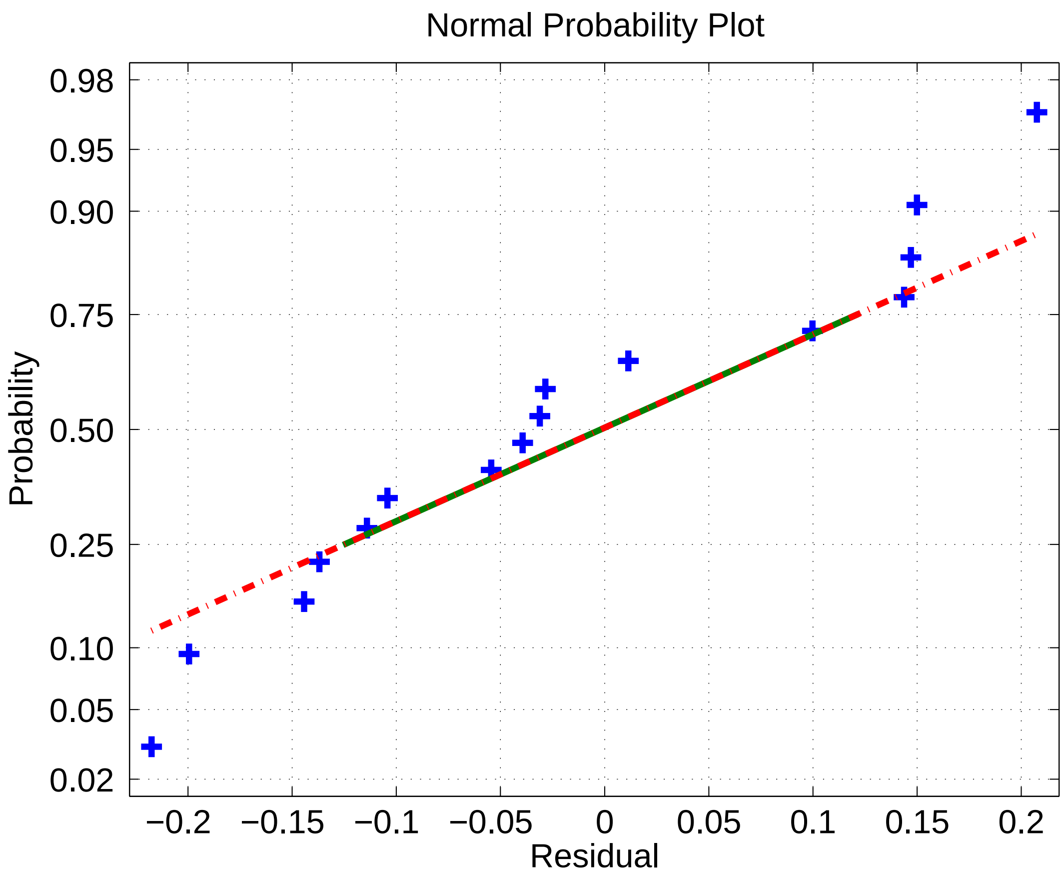
Nonlinear regression residuals



Nonlinear regression residuals



TBS residuals

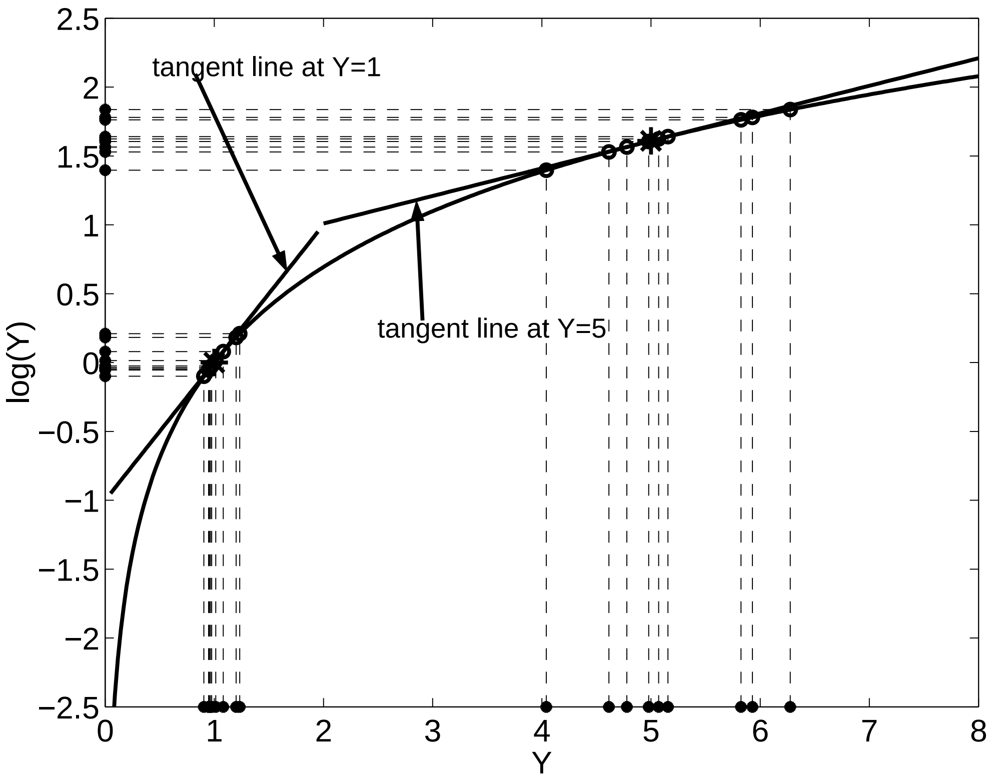


TBS residuals

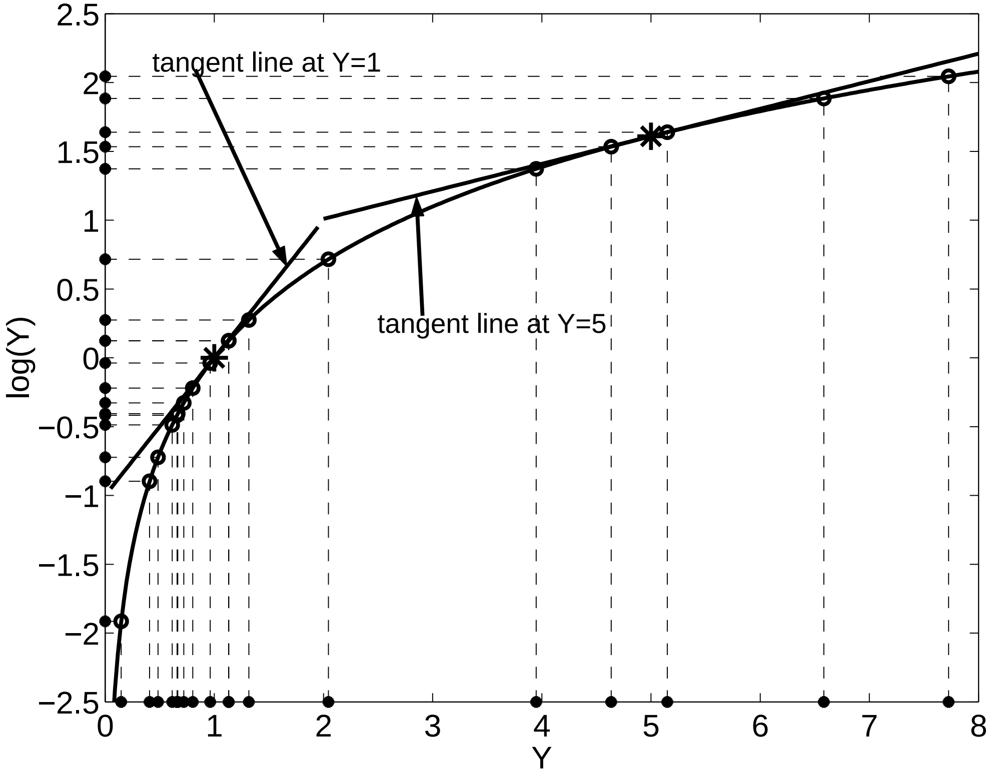
method	$\widehat{Pr}\{\text{default}\vert\text{Aaa}\}$	% of TEXTBOOK estimate
TEXTBOOK	0.005%	100%
nonlinear	0.002%	40%
TBS	0.0008%	16%

Comments:

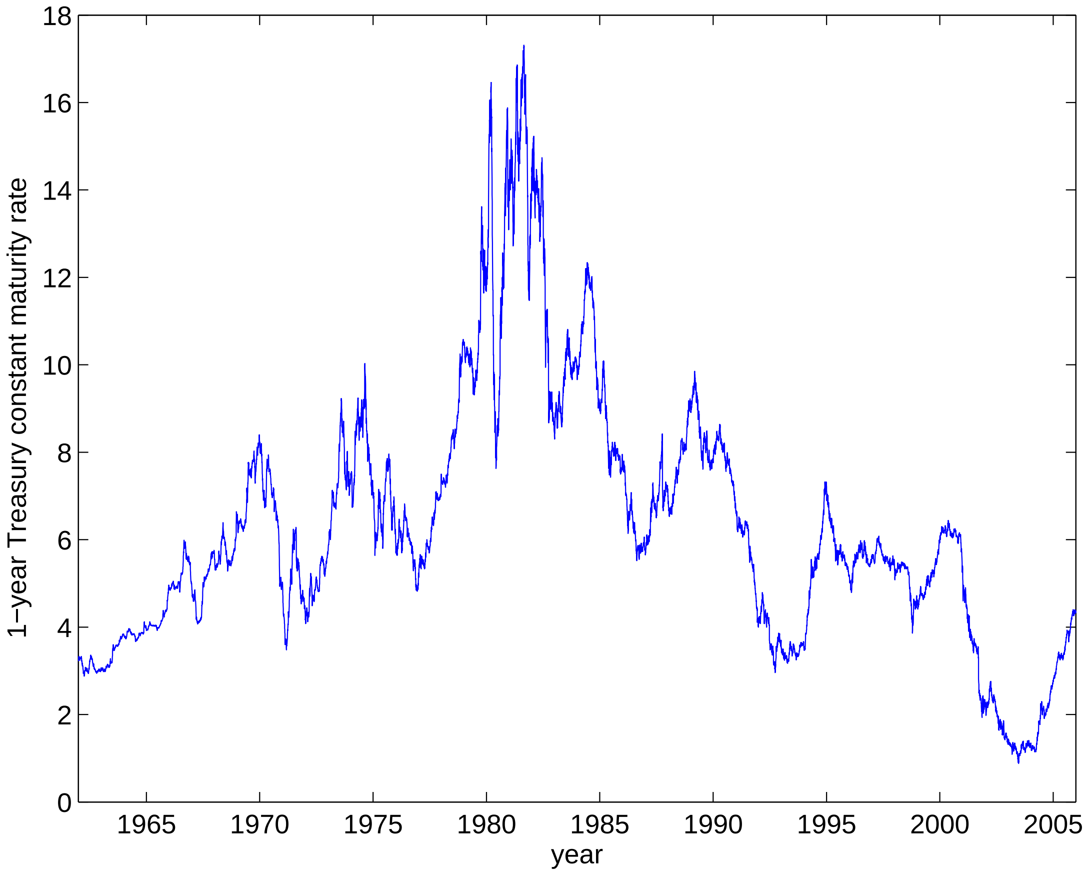
- Suppose sample sizes were large so that all categories had at least one default
  - log transformation would have been applied to all 16 sample proportions
  - but this might have caused outliers and unstable estimates
- Perhaps a logistic regression fit should be compared with the TBS fit.



Geometry of transformations – variance stabilization



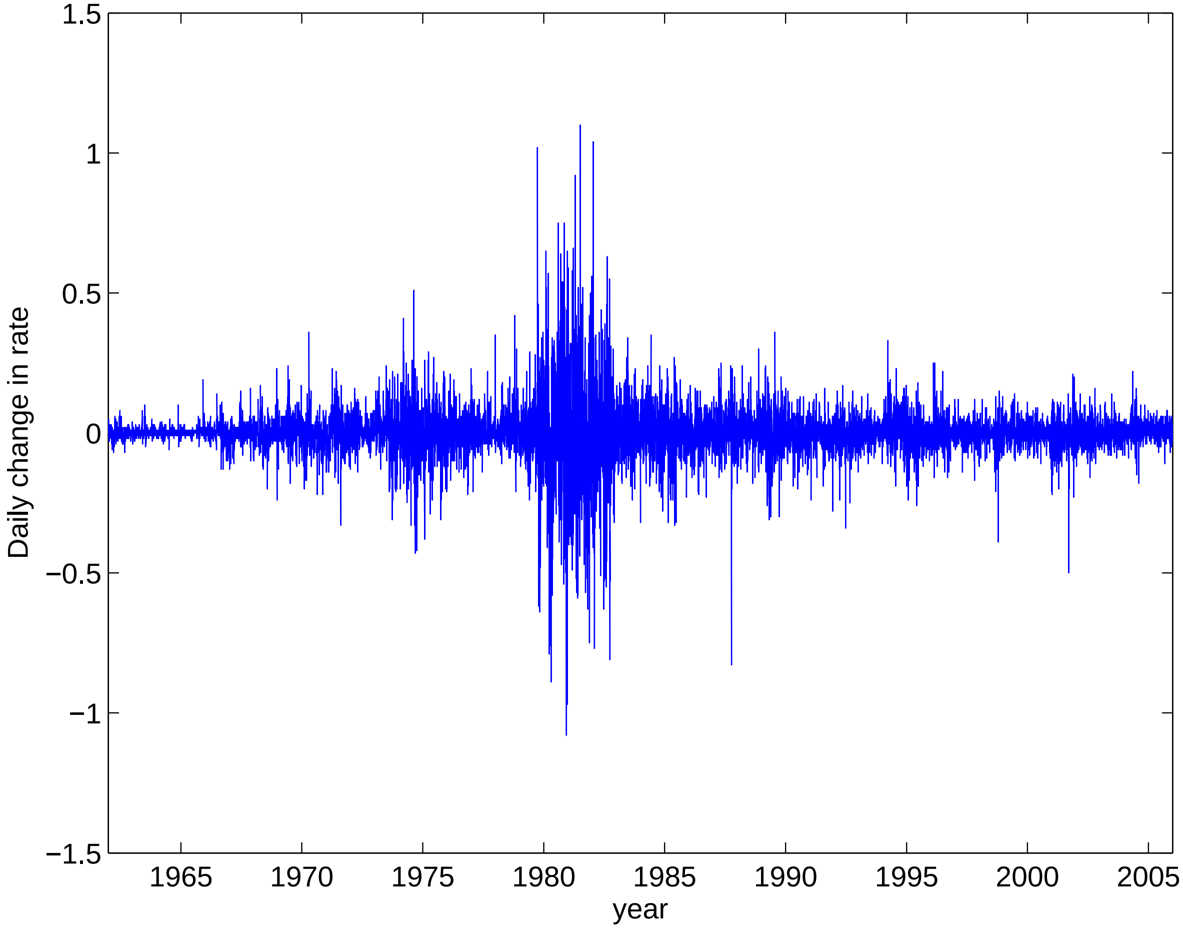
Geometry of transformations – symmetrization



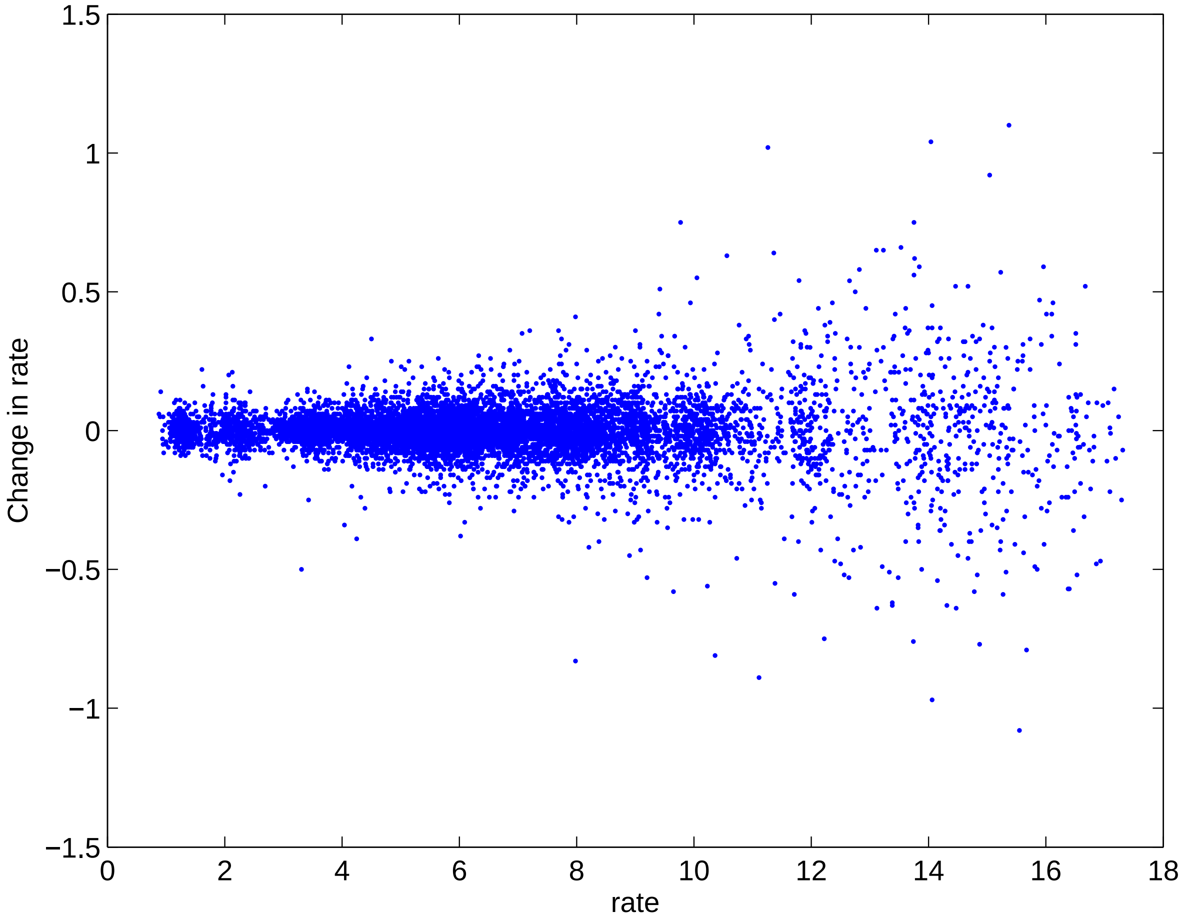
1-Year Treasury Constant Maturity Rate, daily data

Source: Board of Governors of the Federal Reserve System

http://research.stlouisfed.org/fred2/



ΔR_t versus year



ΔR_t versus rate

Estimating Volatility

Parametric model:

$$\mathrm{Var}\{(\Delta R_t)\} = \beta_0 R_{t-1}^{\beta_1}$$

E.g.,

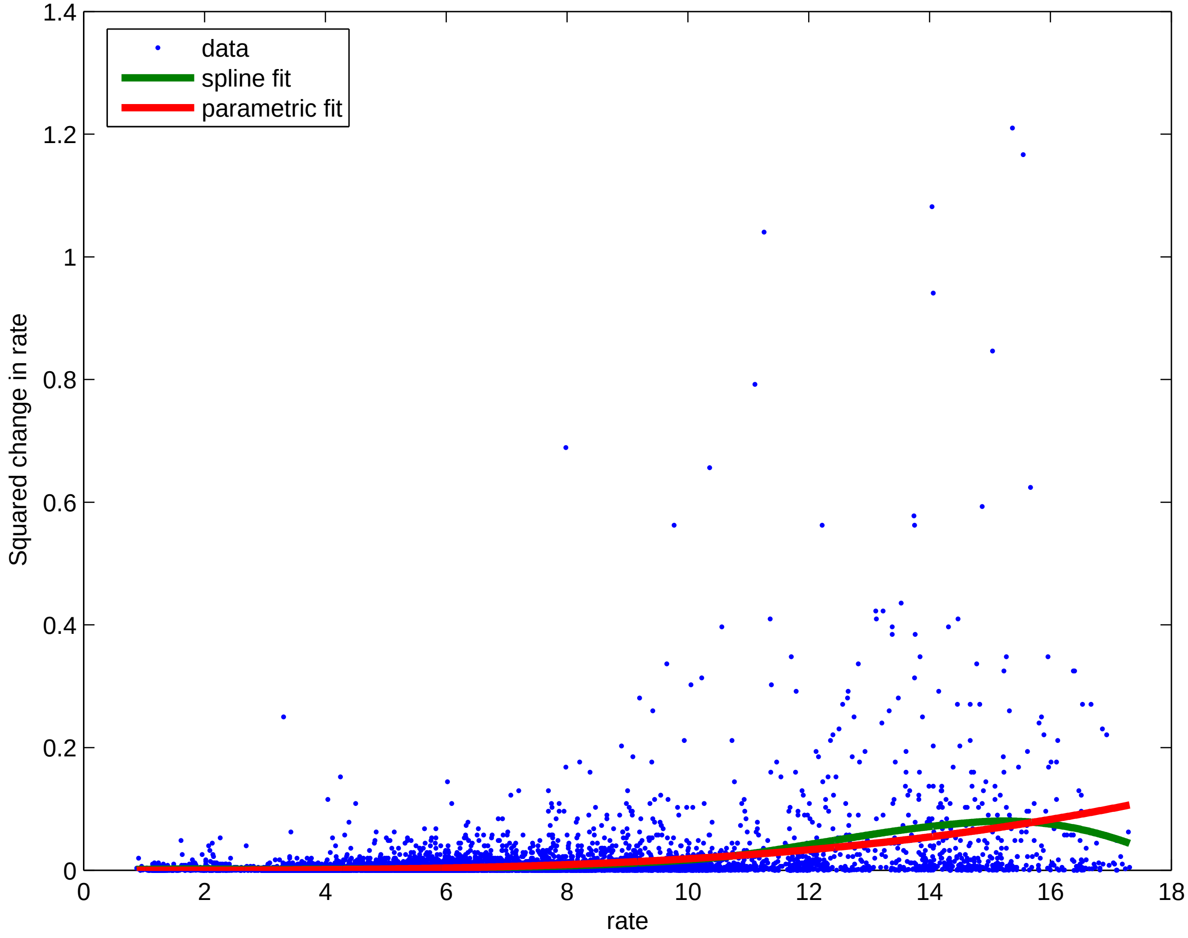
- $\beta_1 = 0$ (Vasicek, 1977)
- $\beta_1 = 1/2$ (Cox, Ingersoll, Ross, 1985)
- $\beta_1 = 1$ (Courtadon, 1982)
- β_1 a free parameter (Chan, Karolyi, Longstaff, and Sanders, 1992)

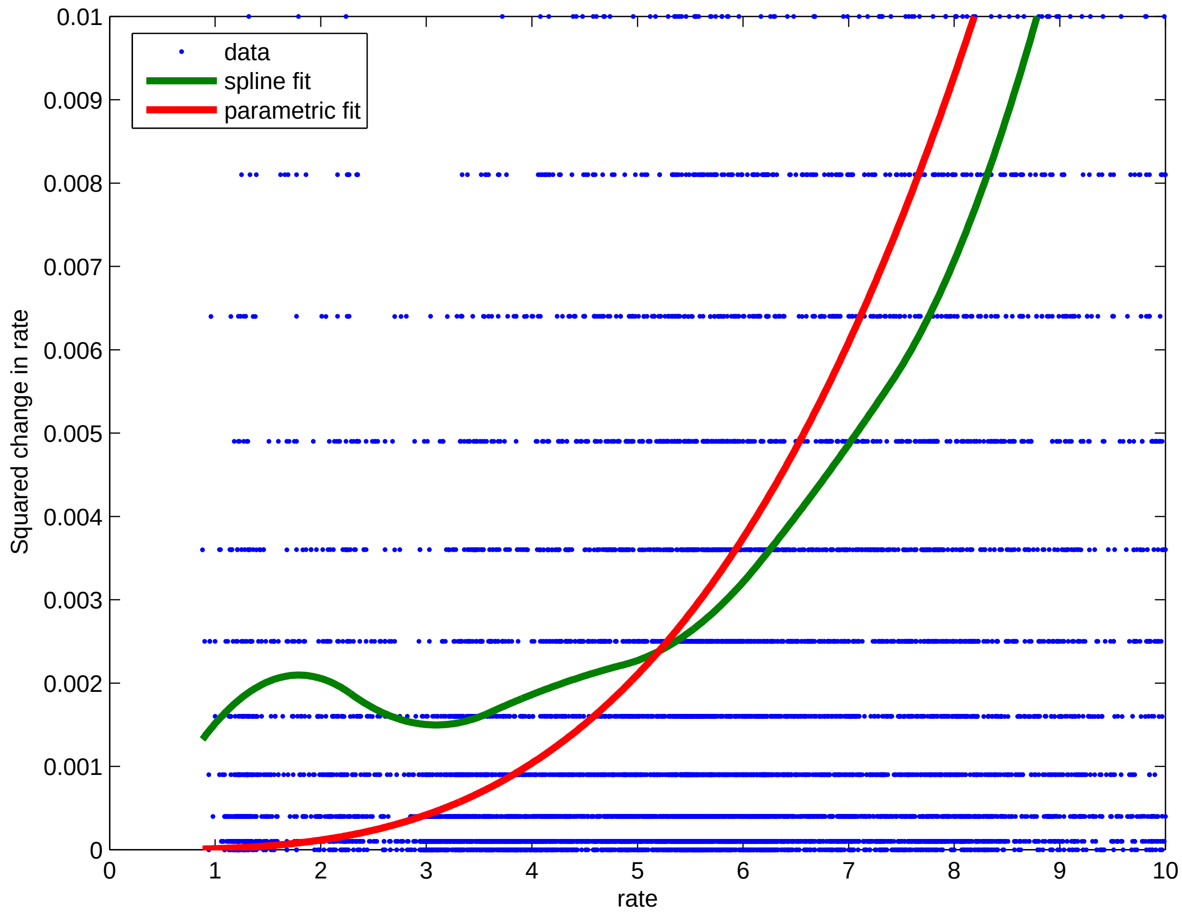
Nonparametric model:

$$\mathrm{Var}\{(\Delta R_t)\} = \sigma^2(R_{t-1})$$

where $\sigma(\cdot)$ is a smooth function

- will be modeled as a spline
- In these models: no dependence on t


data
spline fit
parametric fit
Squared change in rate
rate
0
0.2
0.4
0.6
0.8
1
1.2
1.4
0
2
4
6
8
10
12
14
16
18


data
spline fit
parametric fit
Squared change in rate
rate
0.01
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0
0
1
2
3
4
5
6
7
8
9
10

Penalized Splines for Semiparametric Modeling

Underlying philosophy

1. minimalist statistics
   - keep it as simple as possible
2. build on classical parametric statistics
3. modular methodology

Reference

Semiparametric Regression by Ruppert, Wand, and Carroll (2003)

- Lots of examples.
- But most from biostatistics and epidemiology

Semiparametric regression

Partial linear or partial spline model:

$$Y_i = \mathbf{W}_i^{\mathsf{T}} \boldsymbol{\beta}_W + m(X_i) + \epsilon_i.$$

Here $m(\cdot)$ is a smooth function. We will model it as a spline with a truncated polynomial basis:

$$m(x) = \boldsymbol{X}_i^{\mathsf{T}} \boldsymbol{\beta}_X + \boldsymbol{B}^{\mathsf{T}}(x)\boldsymbol{b}.$$

$$\boldsymbol{X}_i^{\mathsf{T}} = \left(X_i \quad \cdots \quad X_i^p \right)$$

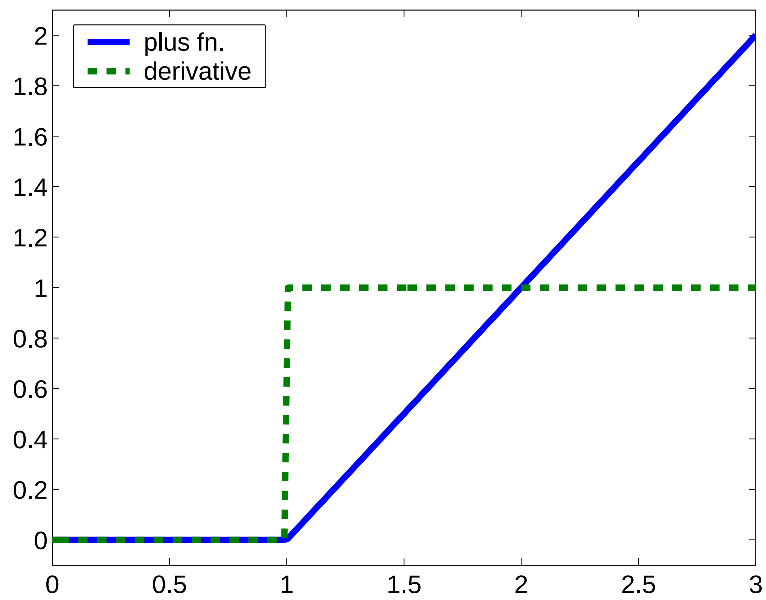
$$\boldsymbol{B}^{\mathsf{T}}(x) = \left\{ (x-\kappa_1)_+^p \quad \cdots \quad (x-\kappa_K)_+^p \right\}$$

The intercept is part of $\mathbf{W}_i^{\mathsf{T}} \boldsymbol{\beta}_W$.

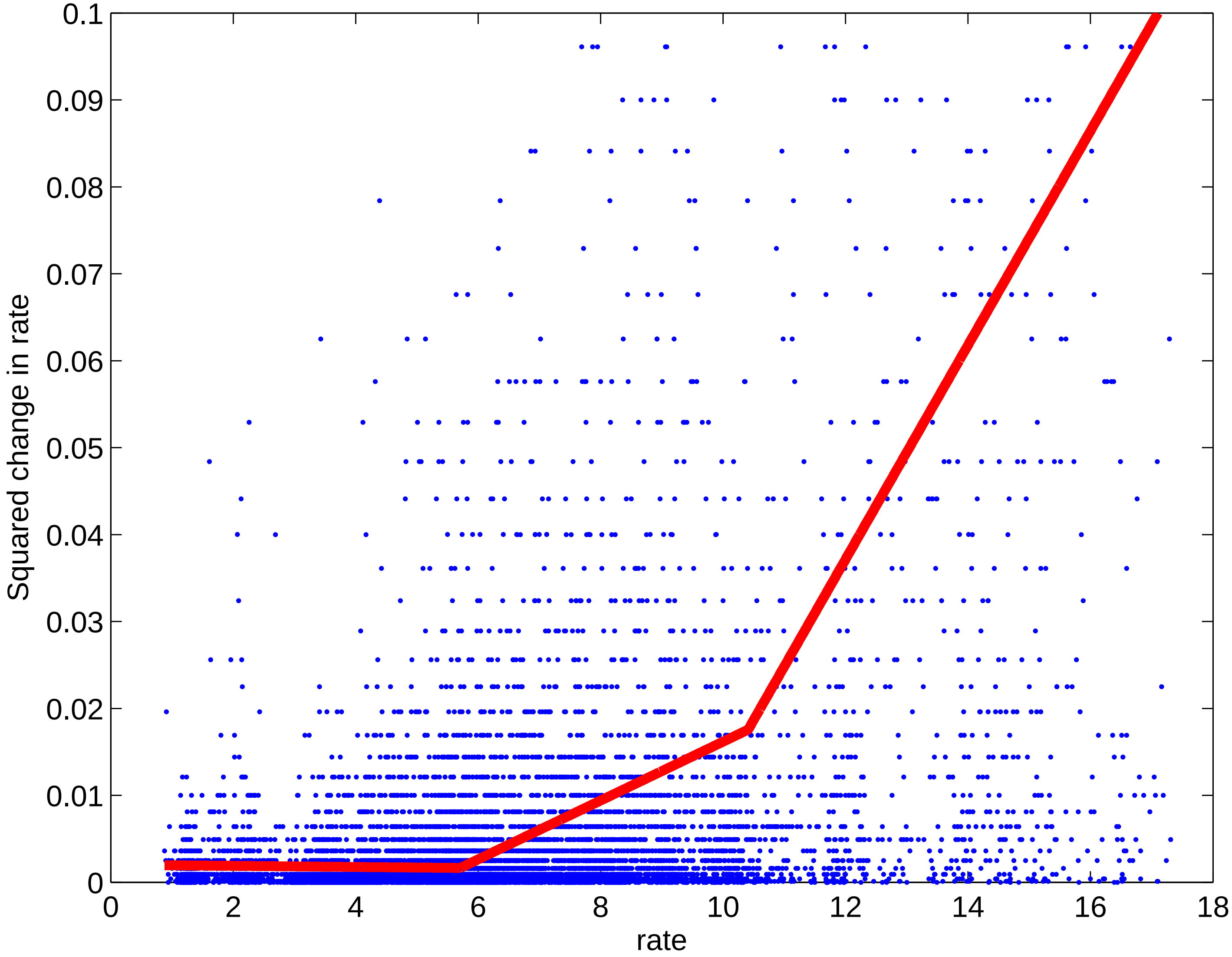
Example

$$m(x) = \beta_1 x + b_1(x - \kappa_1)_+ + \cdots + b_K(x - \kappa_K)_+$$

- slope jumps by b_k at κ_k



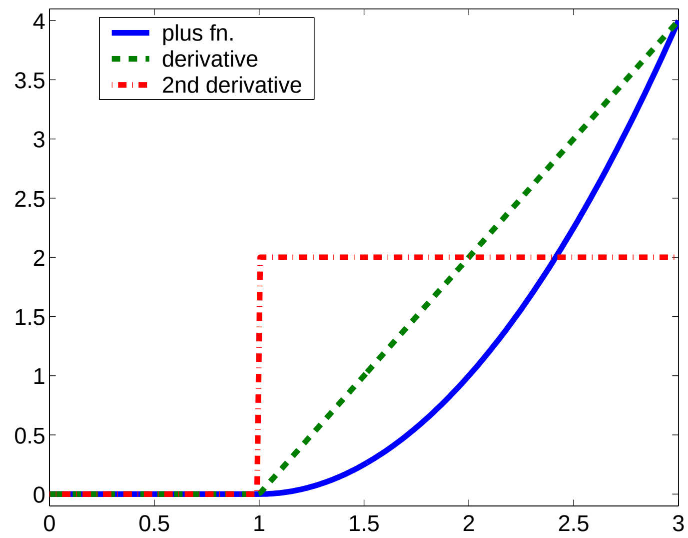
Fitting interest-rate data with plus functions

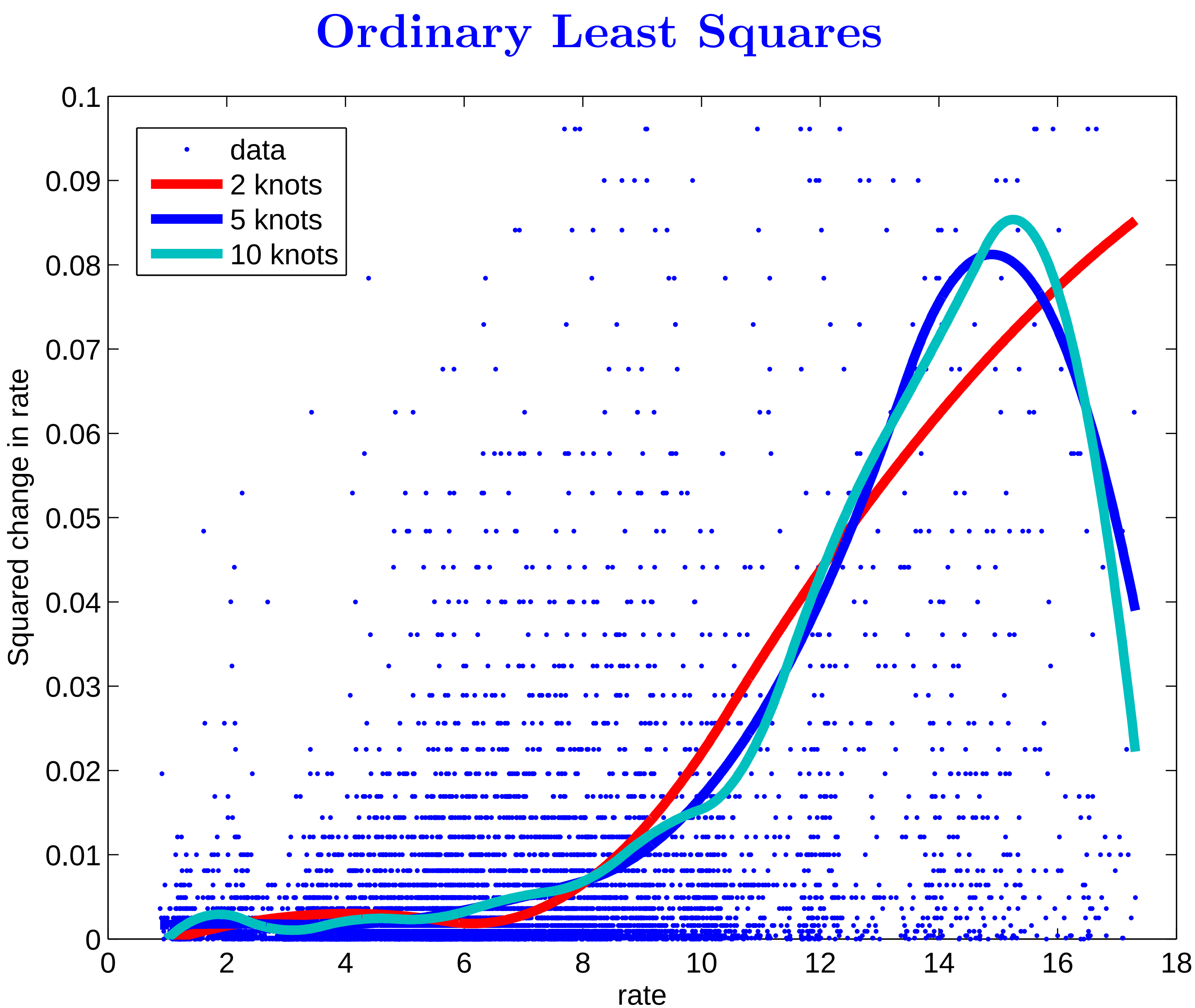


Generalization

$$m(x) = \beta_1 x + \cdots + \beta_p x^p + b_1(x - \kappa_1)_+^p + \cdots + b_K(x - \kappa_K)_+^p$$

- pth derivative jumps by $p!\, b_k$ at κ_k
- first $p - 1$ derivatives are continuous




Ordinary Least Squares
data
2 knots
5 knots
10 knots
Squared change in rate
rate
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0
2
4
6
8
10
12
14
16
18

Penalized least-squares

Minimize

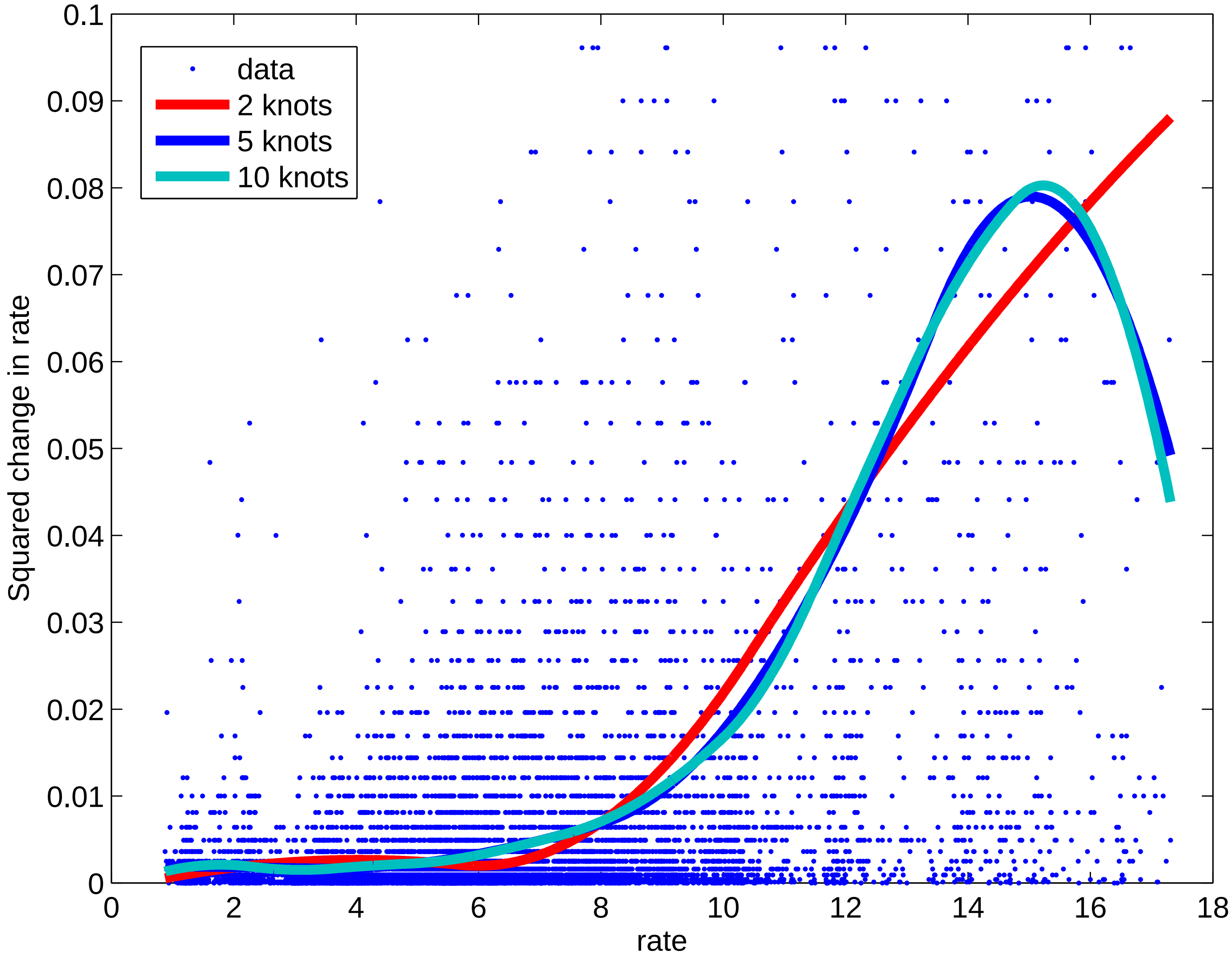
$$\sum_{i=1}^{n} \omega_i^2 \left\{ Y_i - (\mathbf{W}_i^\mathsf{T} \boldsymbol{\beta}_W + \boldsymbol{X}_i^\mathsf{T} \boldsymbol{\beta}_X + \boldsymbol{B}^\mathsf{T}(X_i)\boldsymbol{b}) \right\}^2 + \lambda\, \boldsymbol{b}^\mathsf{T} \boldsymbol{D} \boldsymbol{b}.$$

E.g.,

$$\boldsymbol{D} = \boldsymbol{I}.$$

$$\omega_i = 1/\widehat{\sigma}(Y_i | \mathbf{W}_i, X_i)$$

Penalized Least Squares – Non-adaptive



Ridge Regression

From previous slide:

$$\sum_{i=1}^{n} \omega_i^2 \left\{ Y - (\mathbf{W}_i^\mathsf{T} \boldsymbol{\beta}_W + \boldsymbol{X}_i^\mathsf{T} \boldsymbol{\beta}_X + \boldsymbol{B}^\mathsf{T}(X_i)\boldsymbol{b}) \right\}^2 + \lambda\, \boldsymbol{b}^\mathsf{T} \boldsymbol{D} \boldsymbol{b}.$$

Let $\mathcal{X}$ have row $\left(\mathbf{W}_i^\mathsf{T} \quad \boldsymbol{X}_i^\mathsf{T} \quad \boldsymbol{B}^\mathsf{T}(X_i) \right)$. Then

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}}_W \\ \widehat{\boldsymbol{\beta}}_X \\ \widehat{\boldsymbol{b}} \end{pmatrix} = \left\{ \mathcal{X}^\mathsf{T} \Omega \mathcal{X} + \lambda \text{ blockdiag}(\mathbf{0}, \mathbf{0}, \boldsymbol{D}) \right\}^{-1} \mathcal{X}^\mathsf{T} \Omega \boldsymbol{Y},$$

where

$$\Omega = \text{diag}(\omega_1^2, \ldots, \omega_n^2)$$

Penalized LSE is also

- a BLUP in a mixed model
  - $(\boldsymbol{\beta}_W, \boldsymbol{\beta}_X)$ is the fixed effect vector
  - $\boldsymbol{b}$ is the random effect vector
  - λ is a ratio of variance components
- empirical Bayes estimator.

Selecting λ

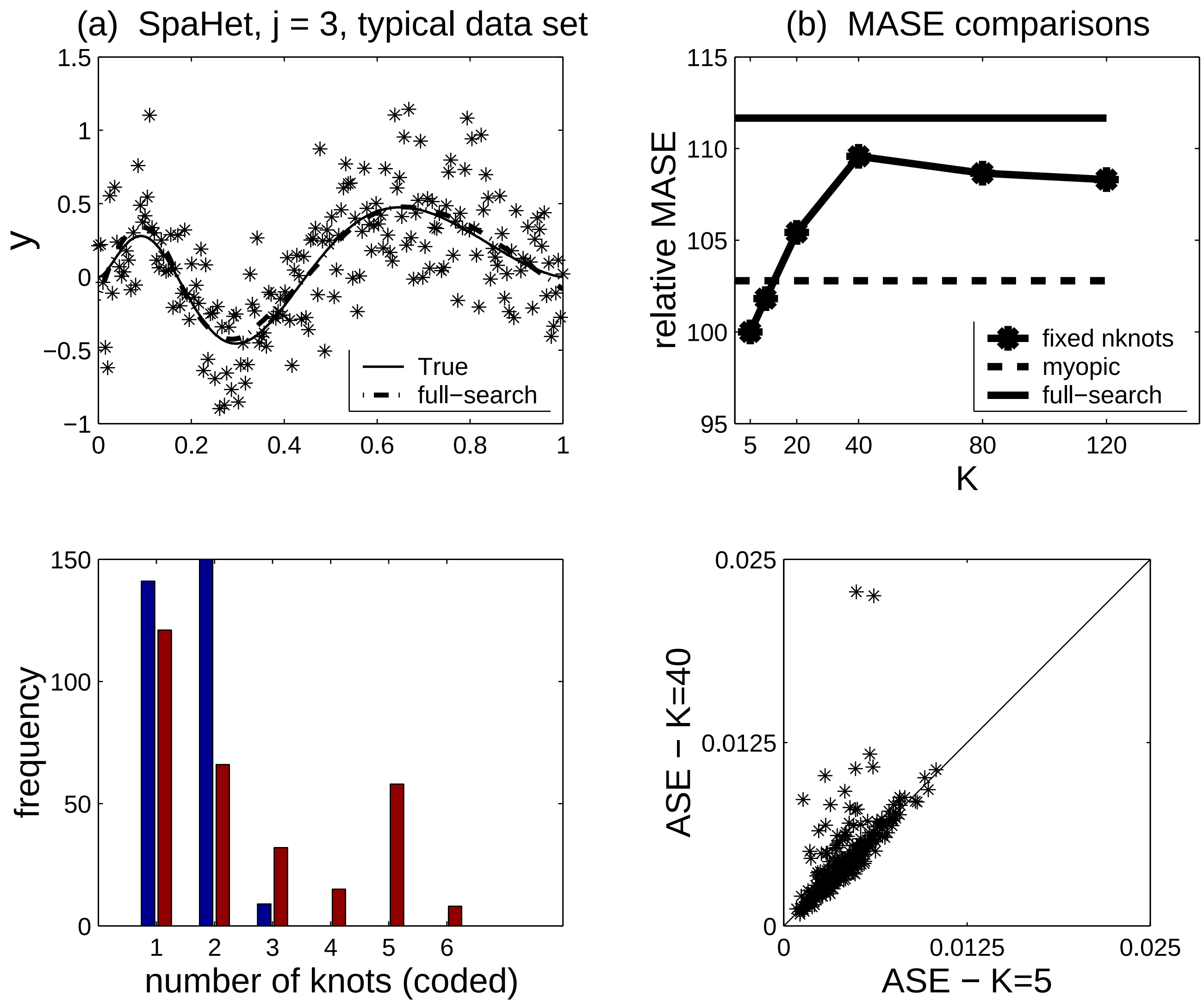
1. cross-validation (CV)
2. generalized cross-validation (GCV)
3. ratio of variance components estimated by ML or REML in mixed model framework
4. as in 3., but estimated in a fully Bayesian framework
5. EBBS = empirical bias bandwidth selection
   - useful if $m'(x)$ is of primary interest

Selecting the Knots Locations

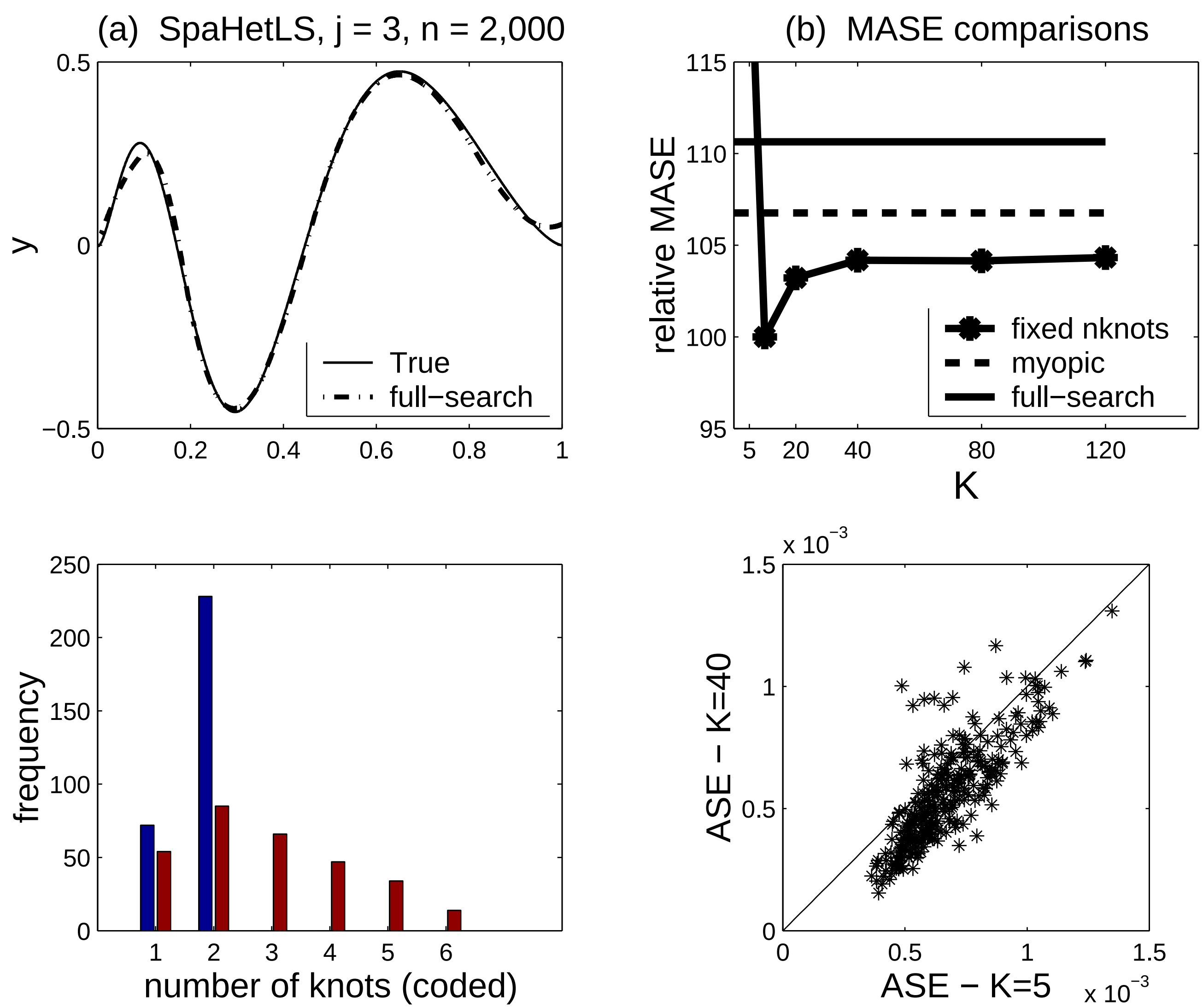
1. I use sample-quantiles of X so there are (approximately) an equal number of observations between any pair of consecutive knots
2. Some prefer equal-spaced knots

1. and 2. give similar results, except in extreme cases.

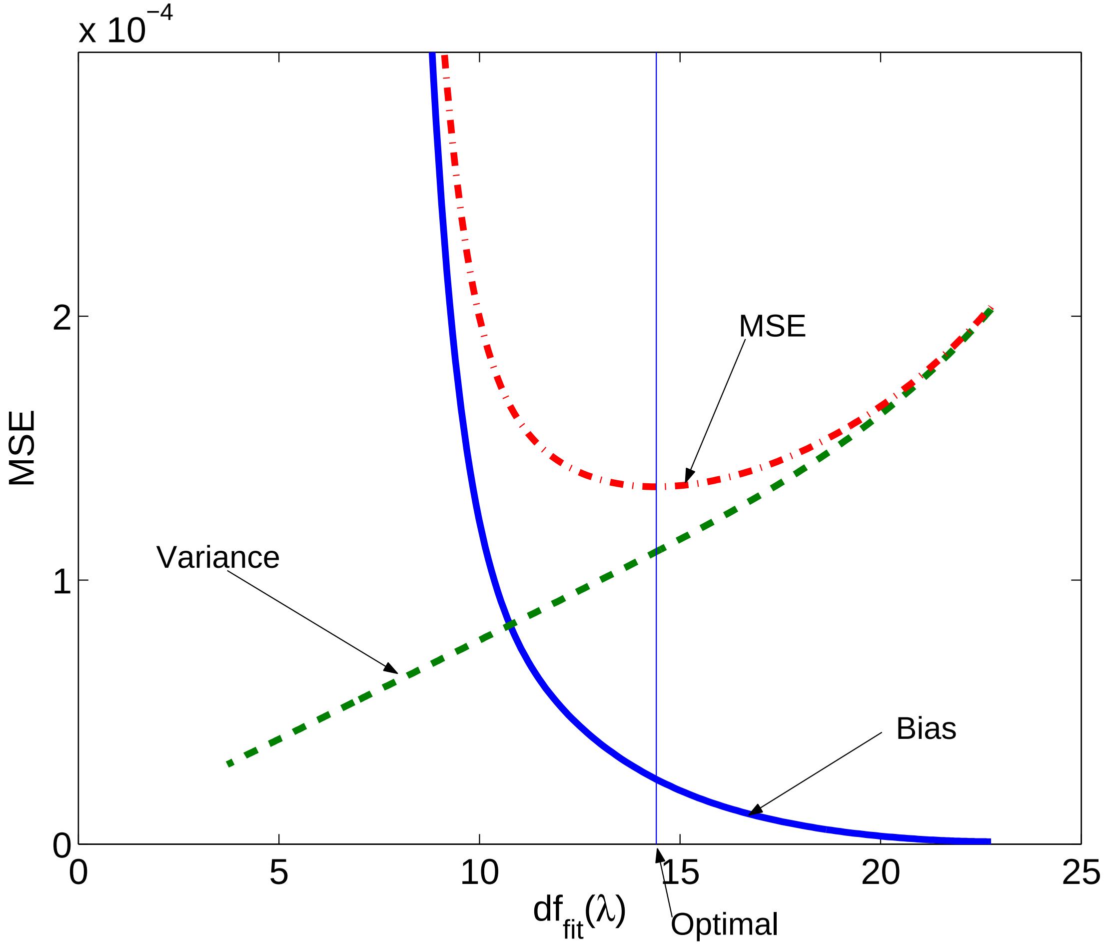
Selecting the Number of Knots



$n = 200$


(a) SpaHetLS, j = 3, n = 2,000
y
True
full−search
(b) MASE comparisons
relative MASE
fixed nknots
myopic
full−search
K
frequency
number of knots (coded)
ASE − K=40
ASE − K=5
x 10^−3

$n = 2,000$



$n = 10,000$, 20 knots, quadratic spline

Additive Models

Model:

$$Y_i = m_1(X_1) + \cdots + m_p(X_p) + \epsilon_i$$

Basis functions:

$$\boldsymbol{X}_{i,j}^{\mathsf{T}} = (\, X_{i,j} \quad \cdots \quad X_{i,j}^{p} \,) \text{ and } \boldsymbol{B}_j^{\mathsf{T}}(x) = \{\, (x - \kappa_{1,j})_+^p \quad \cdots \quad (x - \kappa_{K_j,j})_+^p \,\}$$

Let $\mathcal{X}$ have row

$$(\, \mathbf{W}_i^{\mathsf{T}} \quad \boldsymbol{X}_{i,1}^{\mathsf{T}} \quad \ldots \quad \boldsymbol{X}_{i,p}^{\mathsf{T}} \quad \boldsymbol{B}_1^{\mathsf{T}}(X_{i,1}) \quad \ldots \quad \boldsymbol{B}_p^{\mathsf{T}}(X_{i,p}) \,)$$

Estimation: Minimize

$$\sum_{i=1}^{n} \omega_i^2 \left\{ Y - \left(\mathbf{W}_i^{\mathsf{T}} \boldsymbol{\beta}_W + \sum_{j=1}^{p} \boldsymbol{X}_{i,j}^{\mathsf{T}} \boldsymbol{\beta}_{X,j} + \boldsymbol{B}_j^{\mathsf{T}}(X_{i,j}) \boldsymbol{b}_j \right) \right\}^2 + \sum_{j=1}^{p} \lambda_j \, \boldsymbol{b}_j^{\mathsf{T}} \boldsymbol{D}_j \boldsymbol{b}_j .$$

Adaptive Penalties

- the penalty $\lambda(\cdot)$ is allowed to vary with spatial position
- see Ruppert and Carroll (2000), *Australian and New Zealand Journal of Statistics*
  - $\lambda(\cdot)$ is itself a spline

Minimize:

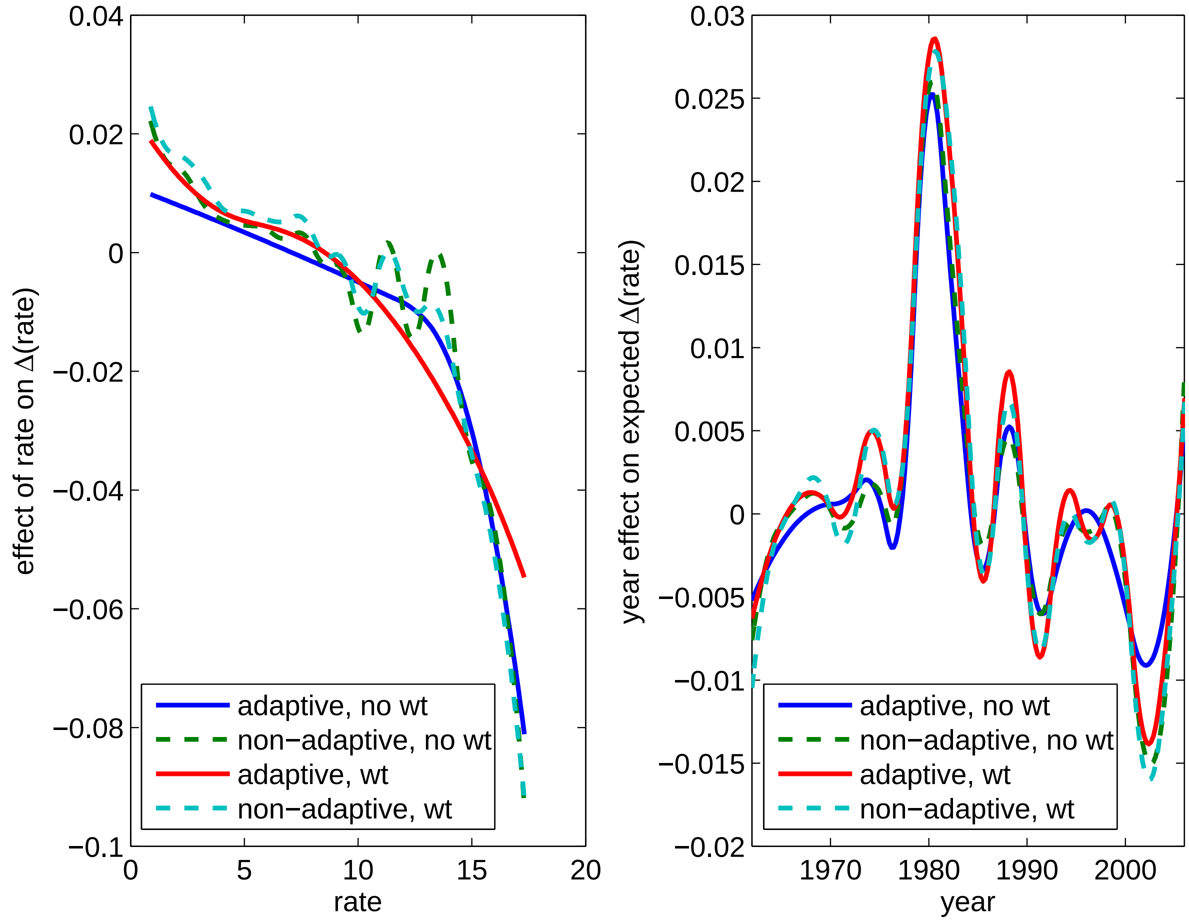
$$\sum_{i=1}^{n} \omega_i^2 \left\{ Y - \left(\mathbf{W}_i^\mathsf{T} \boldsymbol{\beta}_W + \sum_{j=1}^{p} \boldsymbol{X}_{i,j}^\mathsf{T} \boldsymbol{\beta}_{X,j} + \boldsymbol{B}_j^\mathsf{T}(X_{i,j}) \boldsymbol{b}_j \right) \right\}^2 + \sum_{j=1}^{p} \boldsymbol{b}_j^\mathsf{T} \boldsymbol{D}_j \boldsymbol{b}_j.$$

where

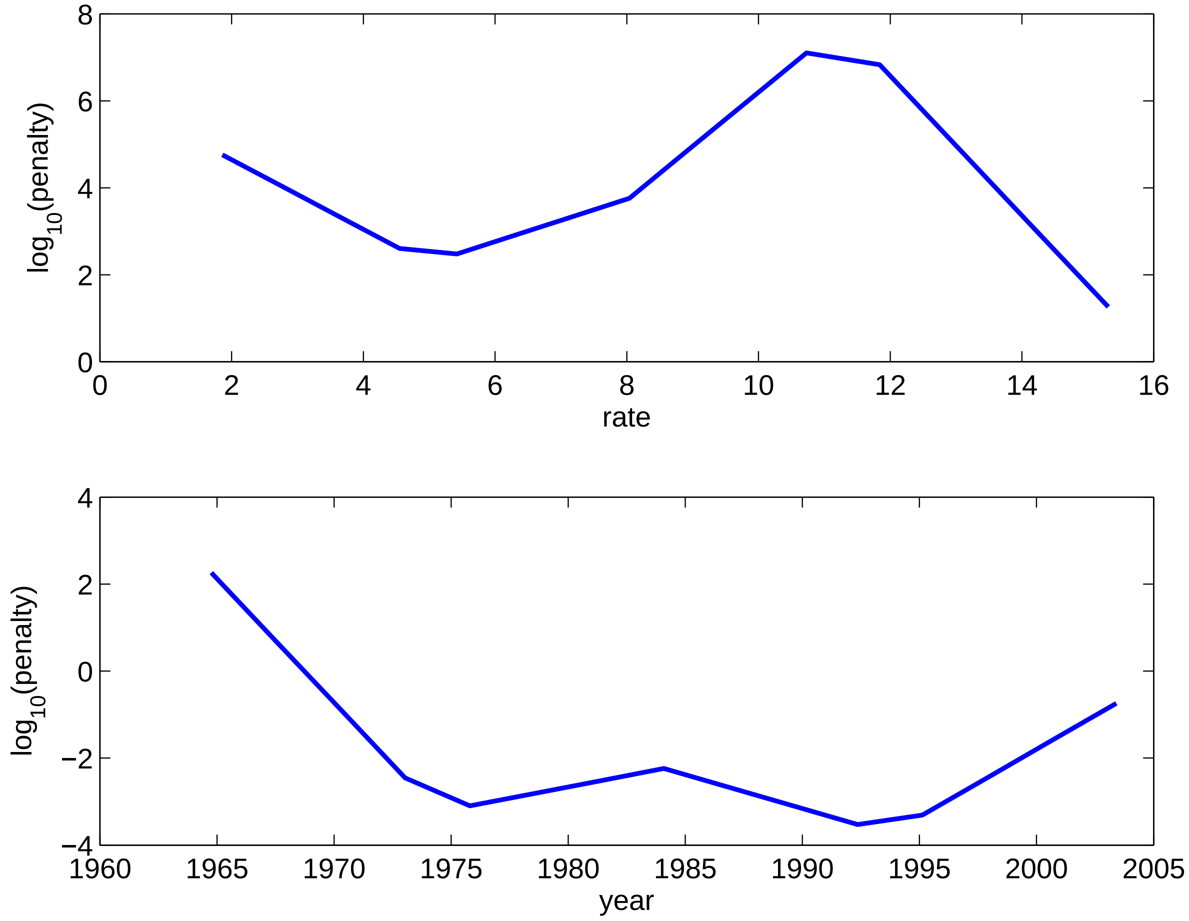
$$\boldsymbol{D}_j = \mathrm{diag}\,(\, \lambda(\kappa_{1,j}) \quad \cdots \quad \lambda(\kappa_{K_j,j})\,)$$

Partial Spline Model

$$\Delta R_t = m_1(R_t) + m_2(t) + \sigma(R_t, t)\epsilon_i$$



Additive fit to ΔR_t



Penalties for adaptive, weighted fit to ΔR_t

Partial Spline Model

$$\Delta R_t = \beta_1 R_t + m_2(t) + \sigma(t, R_t)\epsilon_i$$

Output:

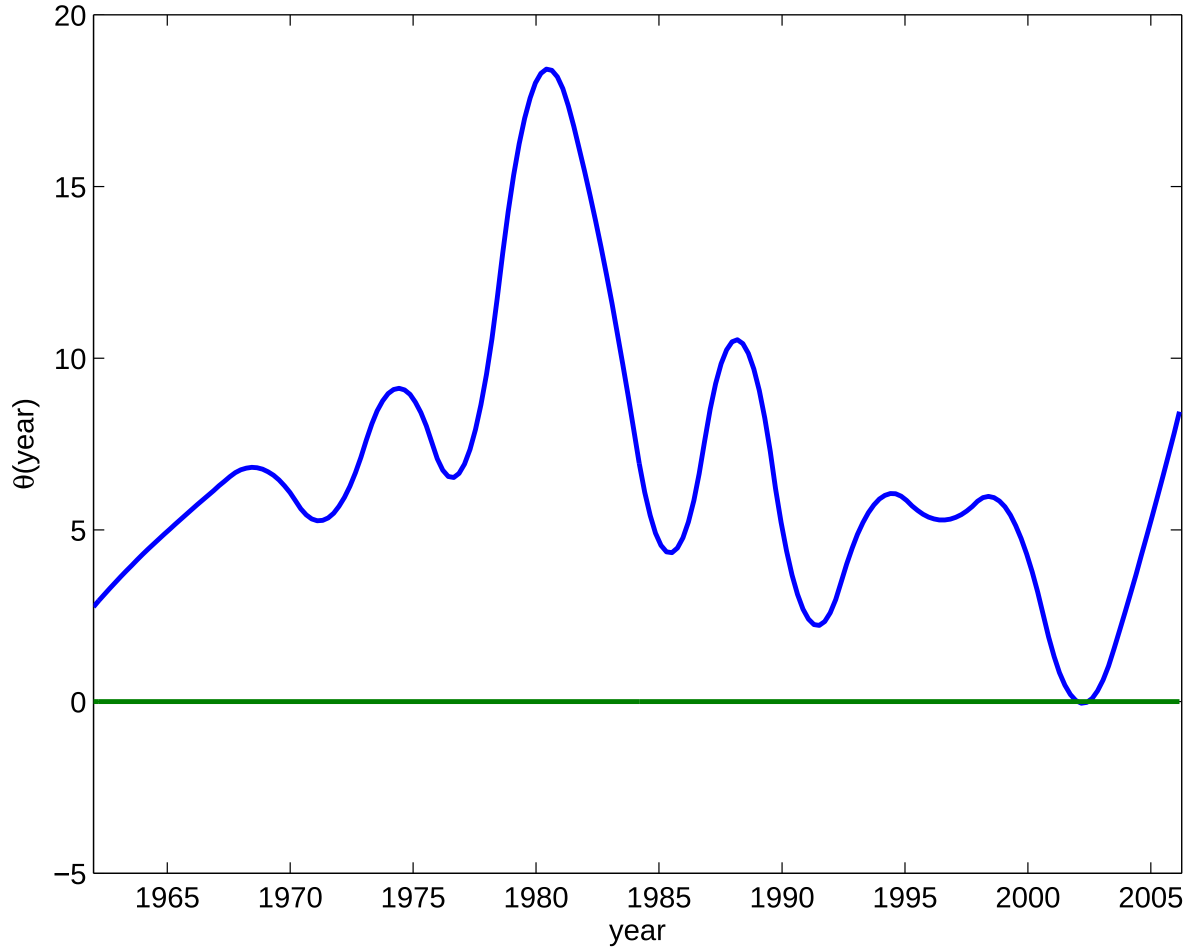
- $\widehat{\beta}_1$
- $\widehat{m}_2(t) + \widehat{\beta}_1 \overline{R_t}$

Corresponds to model with drift:

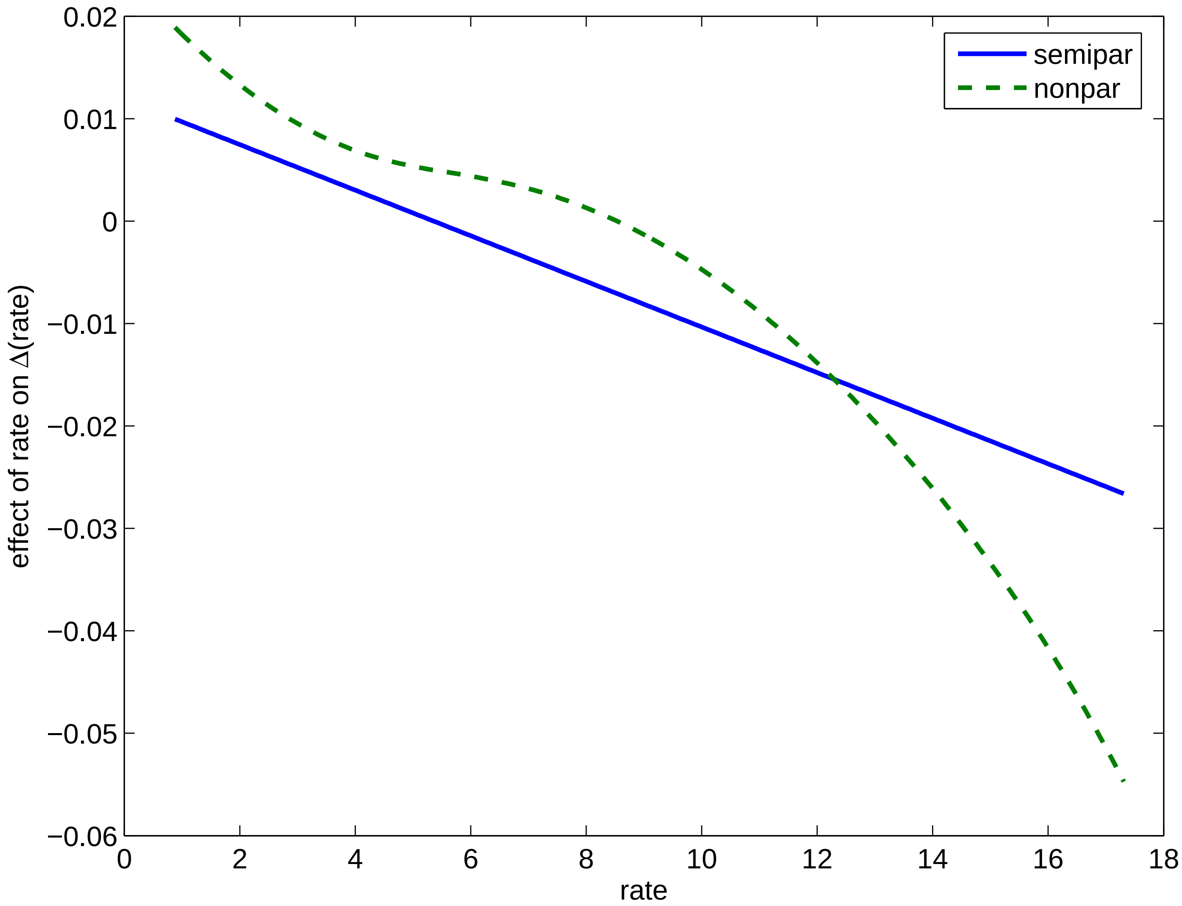
$$a\{\theta(t) - R_t\}$$

where

$$a = -\beta_1 \quad \text{and} \quad \theta(t) = -\frac{m_2(t)}{\beta_1}$$



20
15
10
5
0
−5
θ(year)
1965
1970
1975
1980
1985
1990
1995
2000
2005
year


0.02
0.01
0
−0.01
−0.02
−0.03
−0.04
−0.05
−0.06
effect of rate on Δ(rate)
0
2
4
6
8
10
12
14
16
18
rate
semipar
nonpar

Multiplicative Models for Volatility

$$\mathrm{Var}(Y_t) = \sigma_0^2\, s_1^2(X_1) \,\cdots\, s_p^2(X_p)$$

Example:

$$\mathrm{Var}\{(\Delta R_t)\} = \sigma_0^2 \sigma_1^2(R_{t-1}) \sigma_2^2(t)$$

Backfitting algorithm:

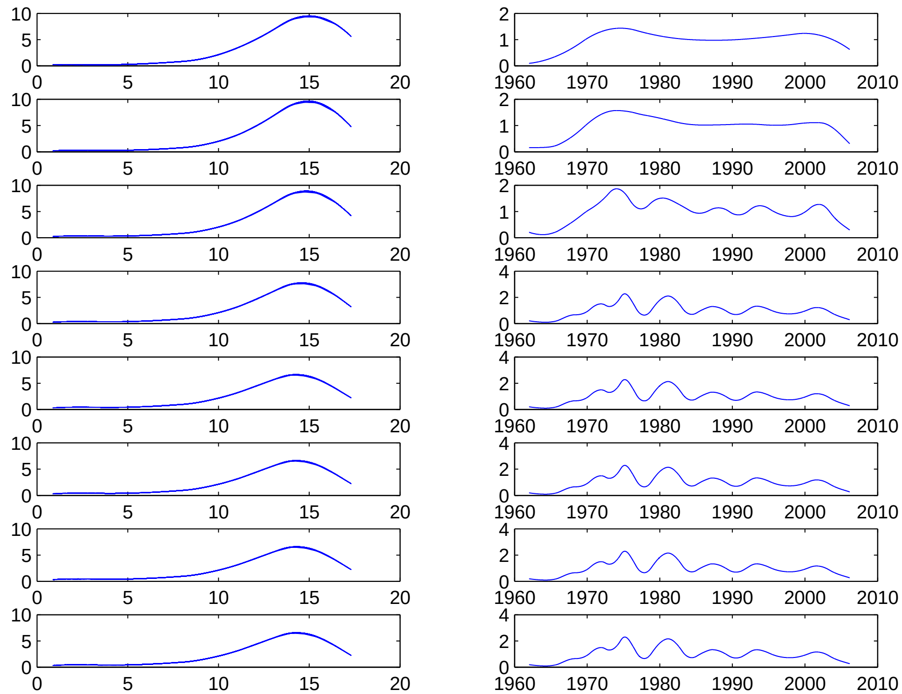
Assume the Y_t has mean zero, e.g., are residuals.

1. fit a model $s_1^2(X_1)$ for Y_t^2 as a function of X_1
   - "de-volatilize": replace Y_t by $y_t/s_1(X_1)$
2. fit a model $s_2^2(X_2)$ for Y_t^2 as a function of X_2
   - "de-volatilize": replace Y_t by $Y_t/s_x(X_2)$

   $\vdots$

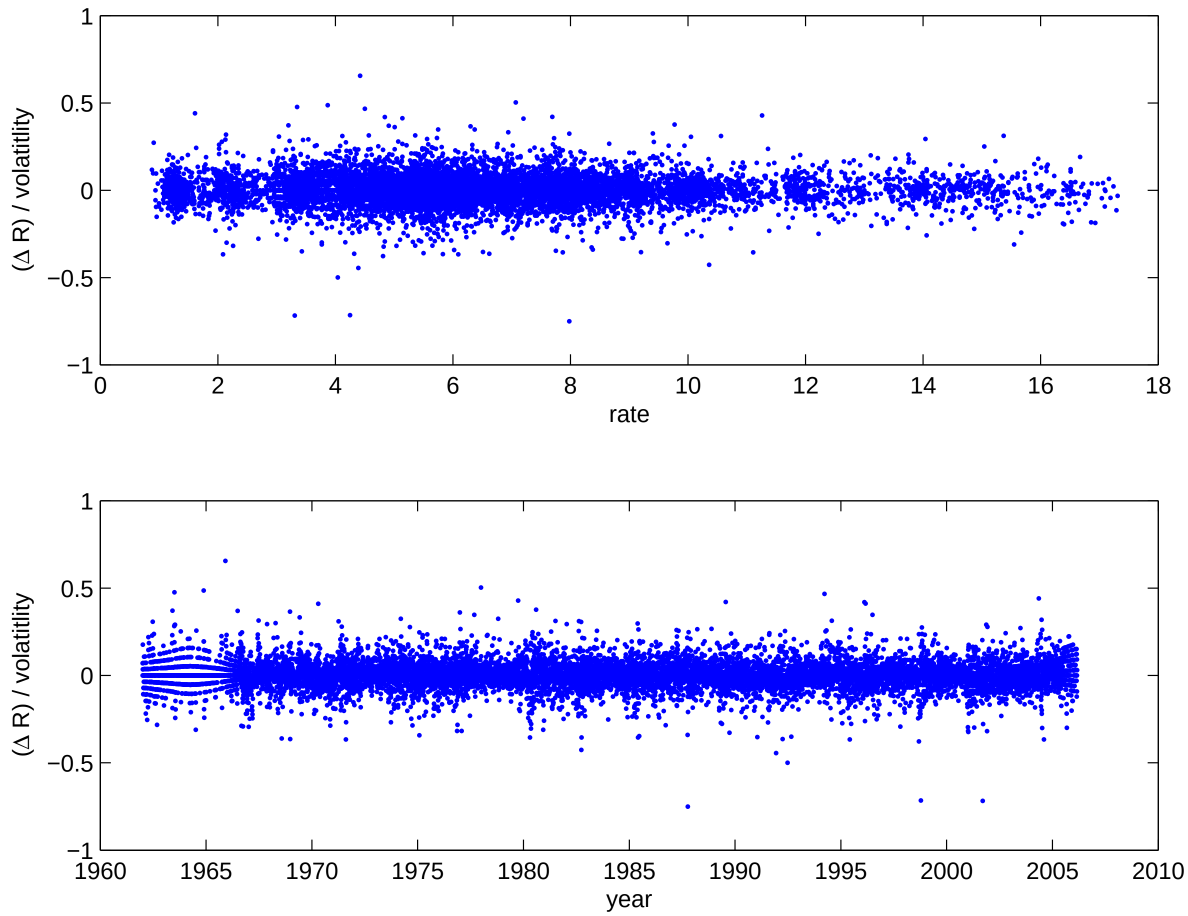
3. fit a model $s_p^2(X_p)$ for Y_t^2 as a function of X_p
   - "de-volatilize": divide Y_t by $s_1(X_1)\cdots s_p(X_p)$
4. either STOP or go back to 1.

Weighting is built into the algorithm.

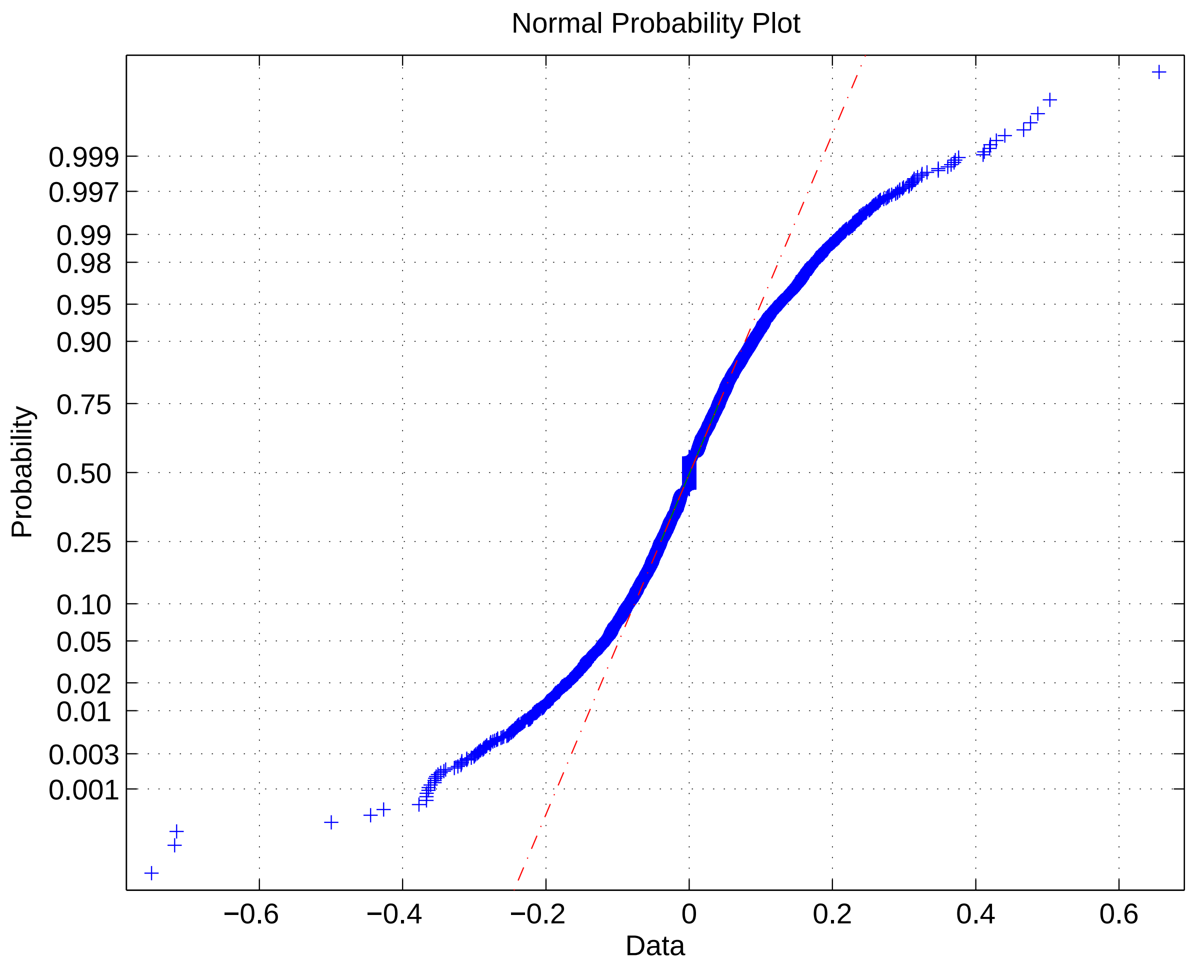


$\{\Delta(R_t)\}^2$ regressed on R_{t-1} and t

- each row is one iteration
- effect of R_{t-1} on left
- effect of t on right
- $\#$ of knots = min(5*iteration number, 20)



Plots of de-volatilized changes versus explanatory variables



Normal plot of de-volatilized changes

Estimating the Term Structure of Corporate Debt with a Semiparametric Model

Joint work with:

- Bob Jarrow (Cornell)
- Yan Yu (University of Cincinnati)

Bond prices and the forward rate

- t = time to maturity
- $P(t)$ = price of zero-coupon bond at current time ($t = 0$)
- $D(t)$ = discount function
- $y(t)$ = yield to maturity
- $f(t)$ = forward rate

$$\frac{P(t)}{\mathrm{PAR}} = D(t) = \exp\{-F(t)\} = \exp\{-ty(t)\} = \exp\left\{-\int_0^t f(s)ds\right\}.$$

Estimation of the forward rate

Suppose the ith bond pays $C_i(t_{i,j})$ and time $t_{i,j}$

- $i = 1, \ldots, n$
- $j = 1, \ldots, z_i$

Let $f(s, \boldsymbol{\delta}) = \boldsymbol{\delta}' \boldsymbol{B}(s)$ be a spline model for the forward rate.

Model for price of ith bond:

$$\widehat{P}_i(\boldsymbol{\delta}) = \sum_{j=1}^{z_i} C_i(t_{i,j}) \exp\{-\boldsymbol{\delta}' \boldsymbol{B}^I(t_{i,j})\}$$

where

$$\boldsymbol{B}^I(t) := \int_0^t \boldsymbol{B}(s) ds = \begin{pmatrix} t & \cdots & \frac{t^{p+1}}{p+1} & \frac{(t-\kappa_1)_+^{p+1}}{p+1} & \cdots & \frac{(t-\kappa_K)_+^{p+1}}{p+1} \end{pmatrix}'.$$

Estimate δ by minimizing

$$Q_{n,\lambda}(\boldsymbol{\delta}) = \frac{1}{n}\sum_{i=1}^{n}\left\{P_i - \sum_{j=1}^{z_i} C_i(t_{i,j})\exp\{-\boldsymbol{\delta}'\boldsymbol{B}^I(t_{i,j})\}\right\}^2 + \lambda\boldsymbol{\delta}'\mathbf{G}\boldsymbol{\delta}.$$

Selection of λ

- Estimation of λ by GCV did not work well
- GCV targets MSE of the estimated regression function
- But the forward rate is the derivative of the (log of) the regression function
- Derivatives require a different amount of smoothing

Corporate Bonds

- Problem: often there are not enough bonds to fit a fully nonparametric model
- Jarrow, Ruppert, and Yu solve this by using a semiparametric model

Algorithm

Step 1: Nonparametric spline fit of a forward rate to US Treasury bonds.

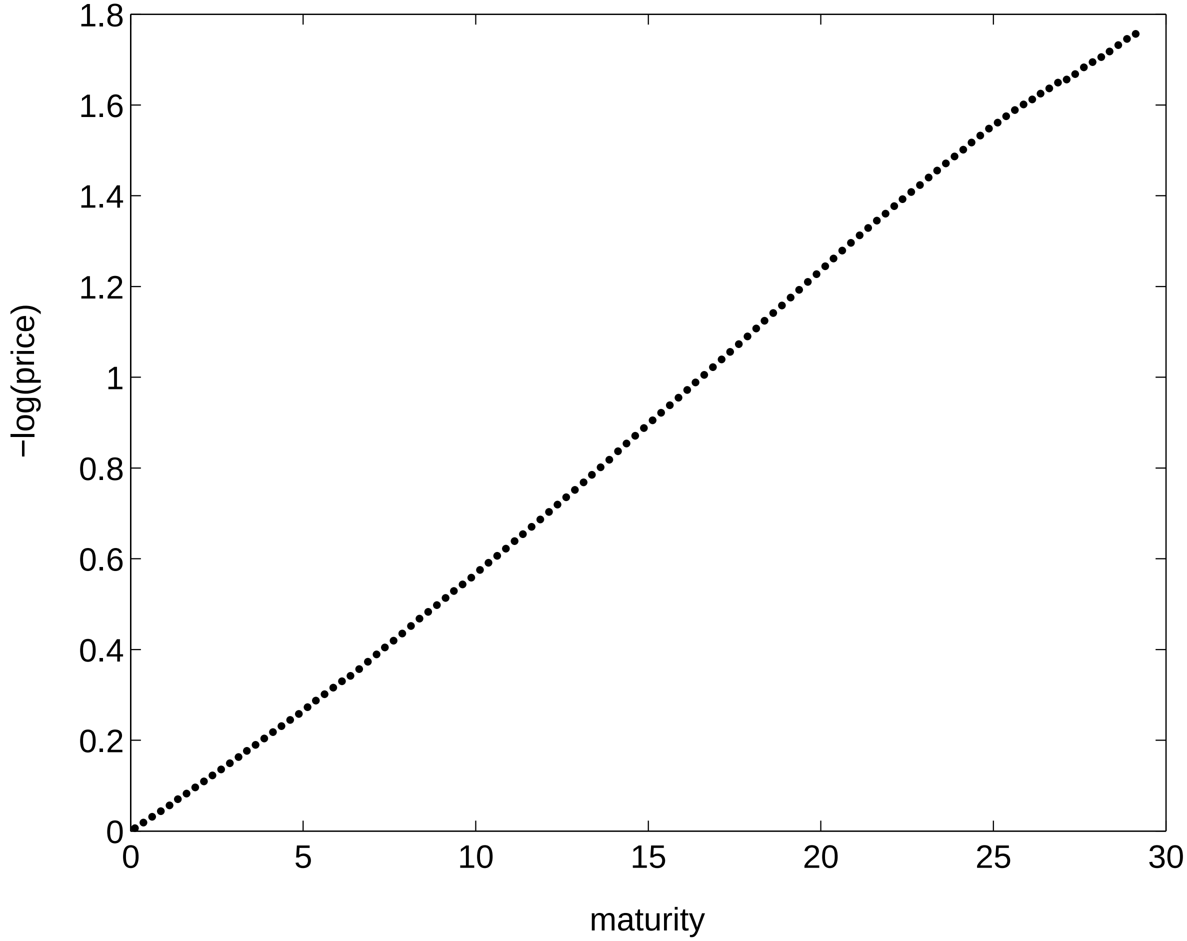
- $\boldsymbol{\delta}$ is estimated by minimizing $Q_{n,\lambda}(\boldsymbol{\delta})$
- λ is chosen by GCV, RSA, or EBBS
- $\widehat{f}_{Tr}(t) = \widehat{\boldsymbol{\delta}}' \boldsymbol{B}(t)$, where $\widehat{\boldsymbol{\delta}}$ are the estimated spline coefficients

Step 2: Parametric estimation to obtain the forward rate curve for a corporation's bonds.

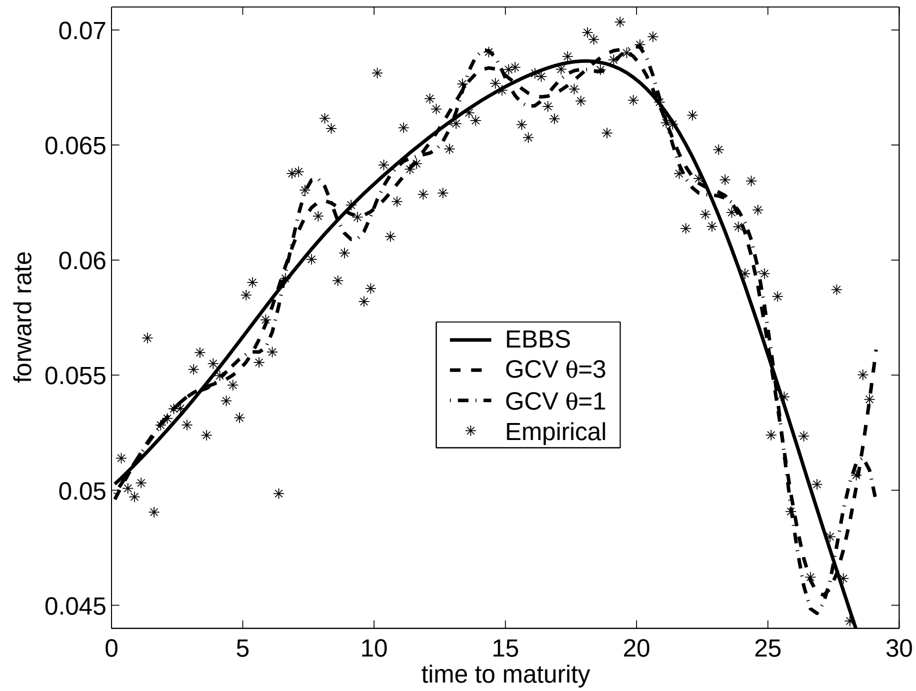
- credit spread is parametric with parameter α
- for example, if the credit spread is a constant, then

$$f_C(t) = \widehat{f}_{Tr}(t) + \alpha = \hat{\boldsymbol{\delta}}' \boldsymbol{B}(t),$$

- fix $\hat{\boldsymbol{\delta}}$ at value from Step 1 and estimate α by OLS



−Log-prices (as fraction of PAR)

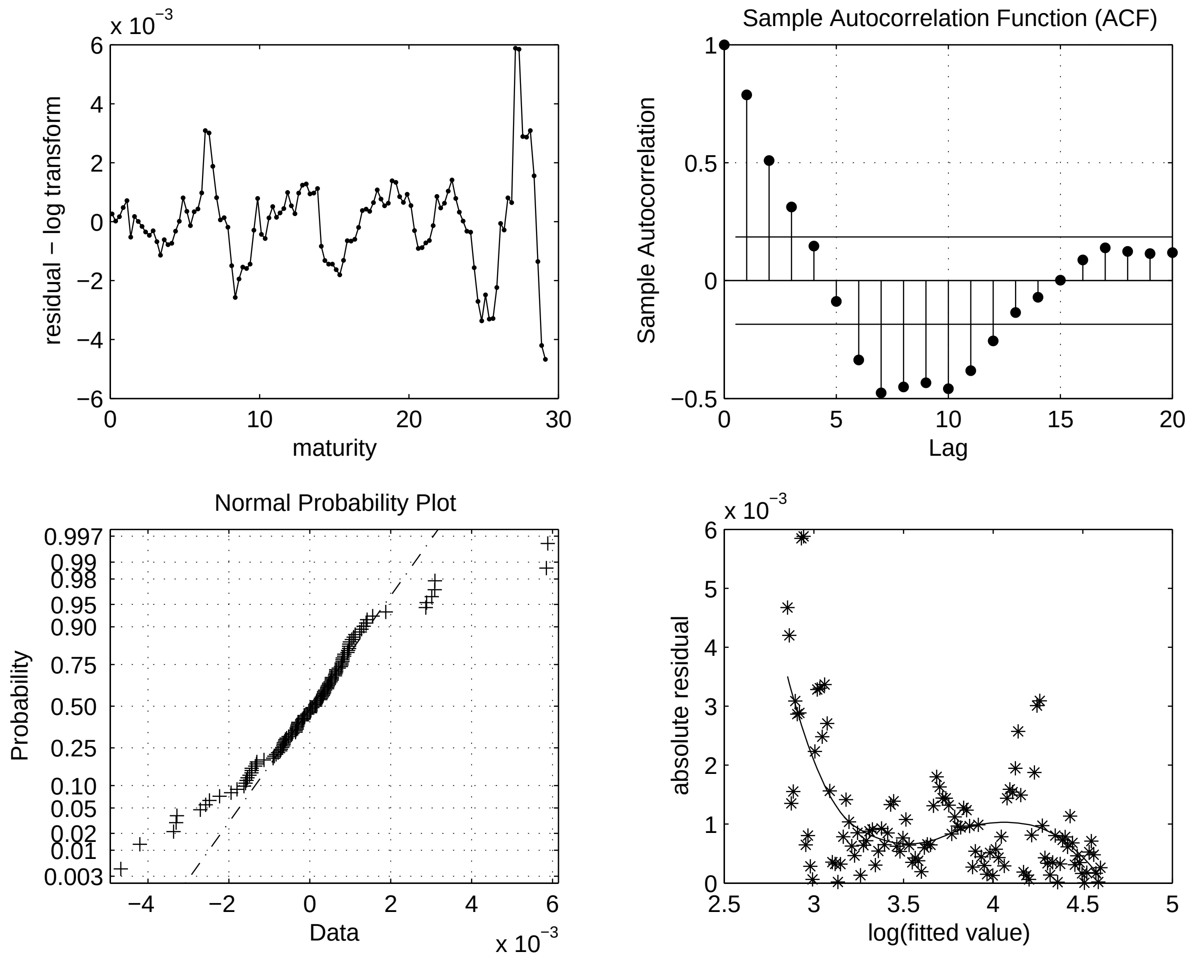


Estimates of forward rate

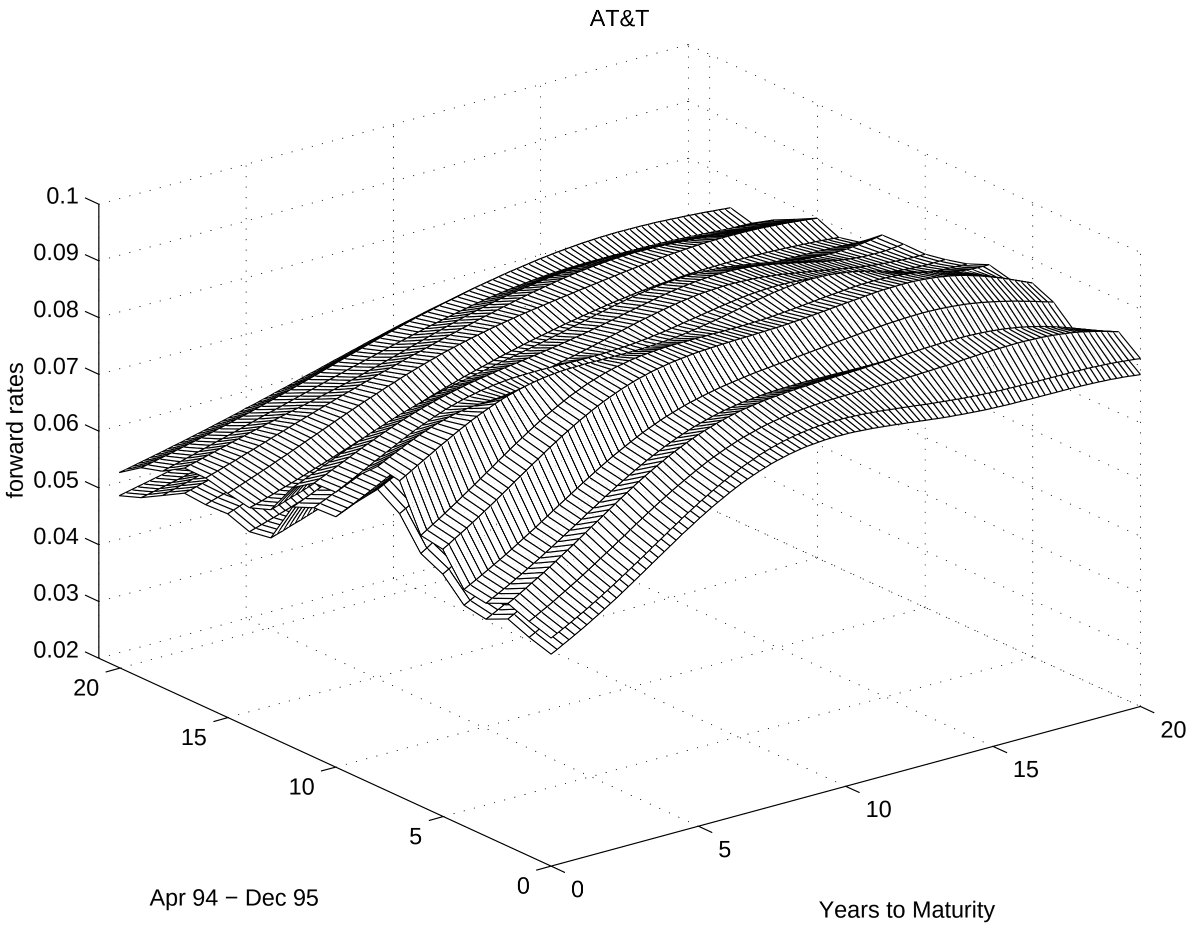
θ was used by Fisher, Nychka and Zervos (1995) to induce more smoothing –

$$GCV(\lambda) = \frac{n^{-1}\sum_{i=1}^{n}\left\{P_i - \widehat{P}_i(\boldsymbol{\delta})\right\}^2}{\{1 - n^{-1}\theta\,\mathrm{tr}\boldsymbol{A}(\lambda)\}^2},$$

where $\boldsymbol{A}(\lambda)$ is the "hat" or "smoother" matrix: $\widehat{\mathbf{P}} = \boldsymbol{A}(\lambda)\mathbf{P}$



Residual analysis



Estimates of Treasury and AT&T forward rates

Summary

- Statisticians and financial engineers would each benefit from more collaboration
- Calibration of financial models is an interesting and challenging problem in statistics and data analysis
  - transformation and weighting can be important
- Penalized splines are an attractive method for semiparametric modeling