Estimating Term Structure with Penalized Splines

David Ruppert
Operations Research and Industrial Engineering
Cornell University

Joint work with Robert Jarrow and Yan Yu

January 27, 2004

Outline

- Bond prices, forward rates, yields
- Empirical forward rate – noisy
- Modelling the forward rate
- Penalized least-squares
- Inadequacy of cross-validation
- Residual analysis – checking the noise assumptions
- Corporate term structure and credit spreads
- Asymptotics

Discount Function, Forward Rates, and Yields

- $D(0,t) = D(t)$ is the discount function, the value at time 0 (now) of a zero-coupon bond that pays \$1 at time t.

$$\frac{\text{Price}(t)}{\text{PAR}} = D(t)$$

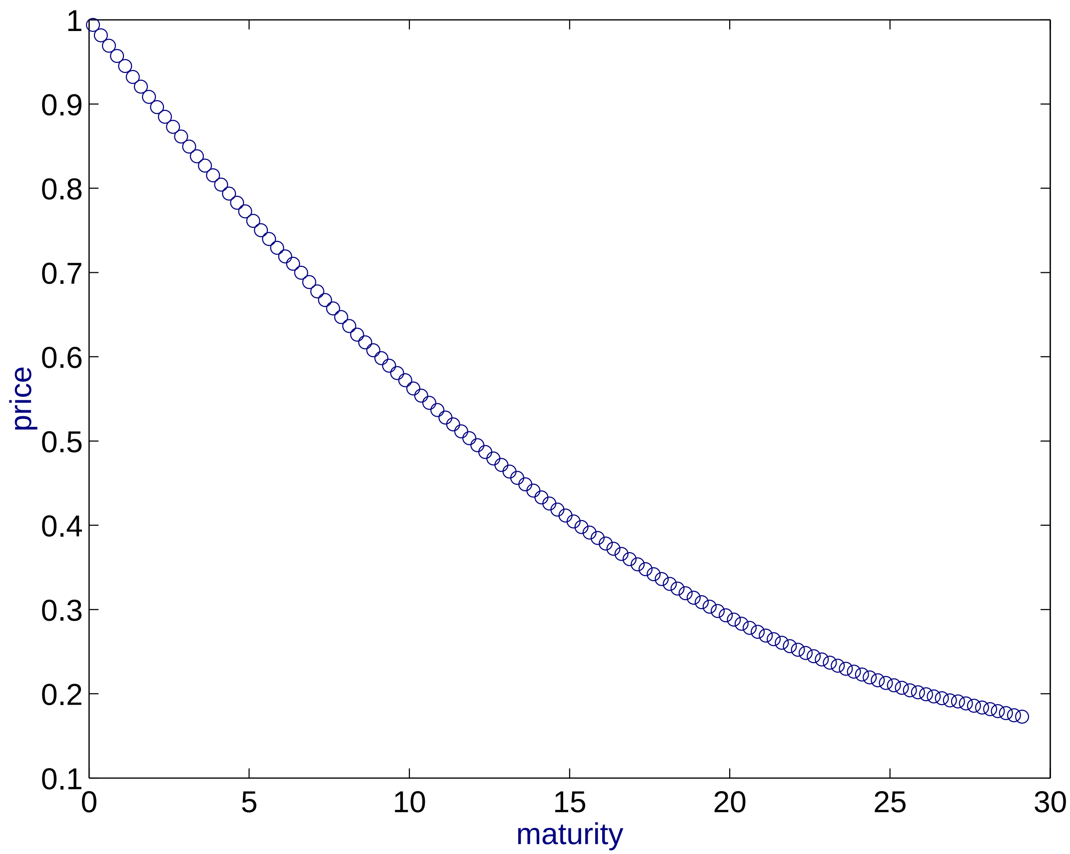
- $f(t)$ is the current forward rate defined by

$$D(t) = \exp\left\{-\int_0^t f(s)ds\right\} \text{ for all } t$$

- The yield is the average forward rate, i.e.,

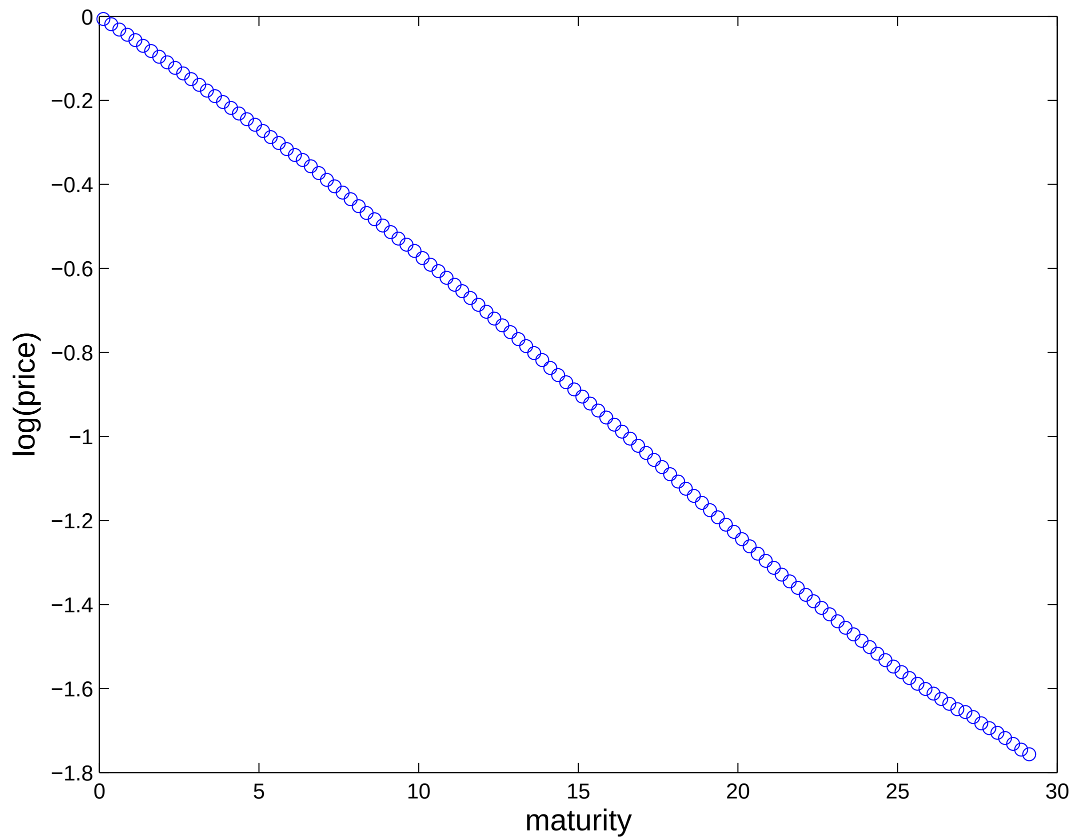
$$y(t) = \frac{1}{t}\int_0^t f(s)ds = -\frac{1}{t}\log\{D(t)\}$$

Discount Function, Forward Rates, and Yields



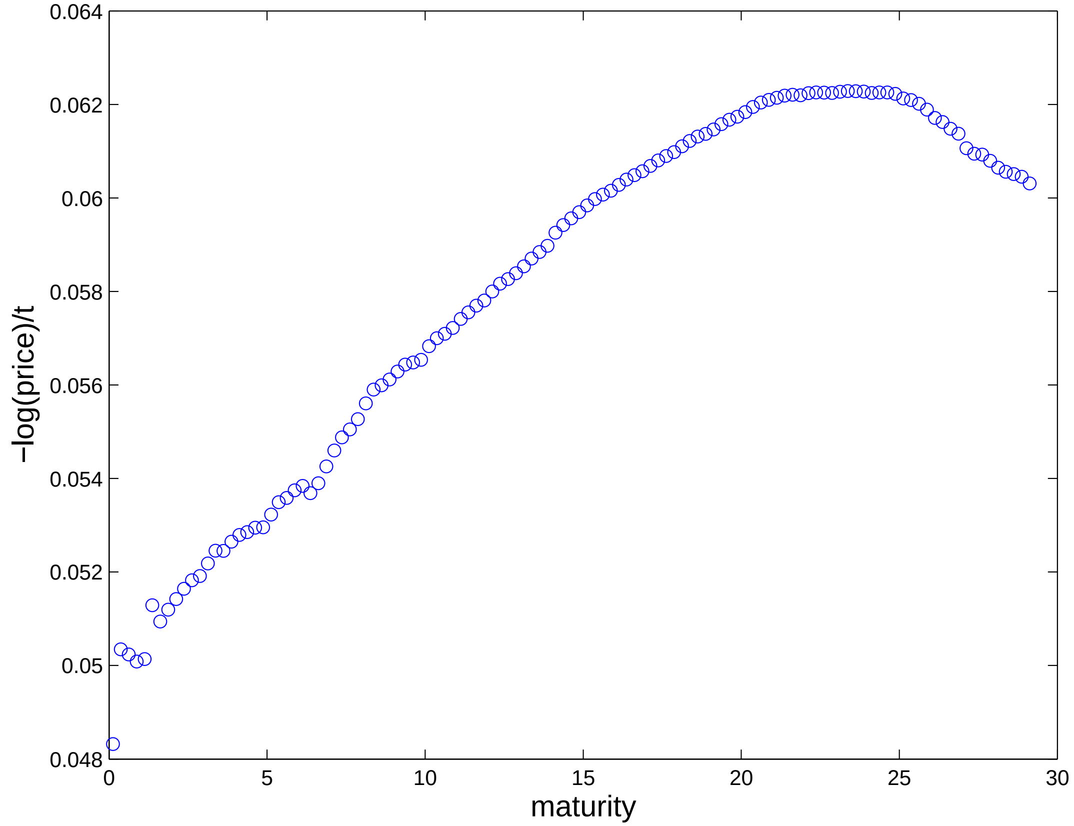
STRIPS on Dec 31, 1995: price $=$ empirical discount function

Prices, Forward Rates, and Yields



STRIPS on Dec 31, 1995: log prices

Discount Function, Forward Rates, and Yields



STRIPS on Dec 31, 1995: empirical yields

Empirical Forward Rate

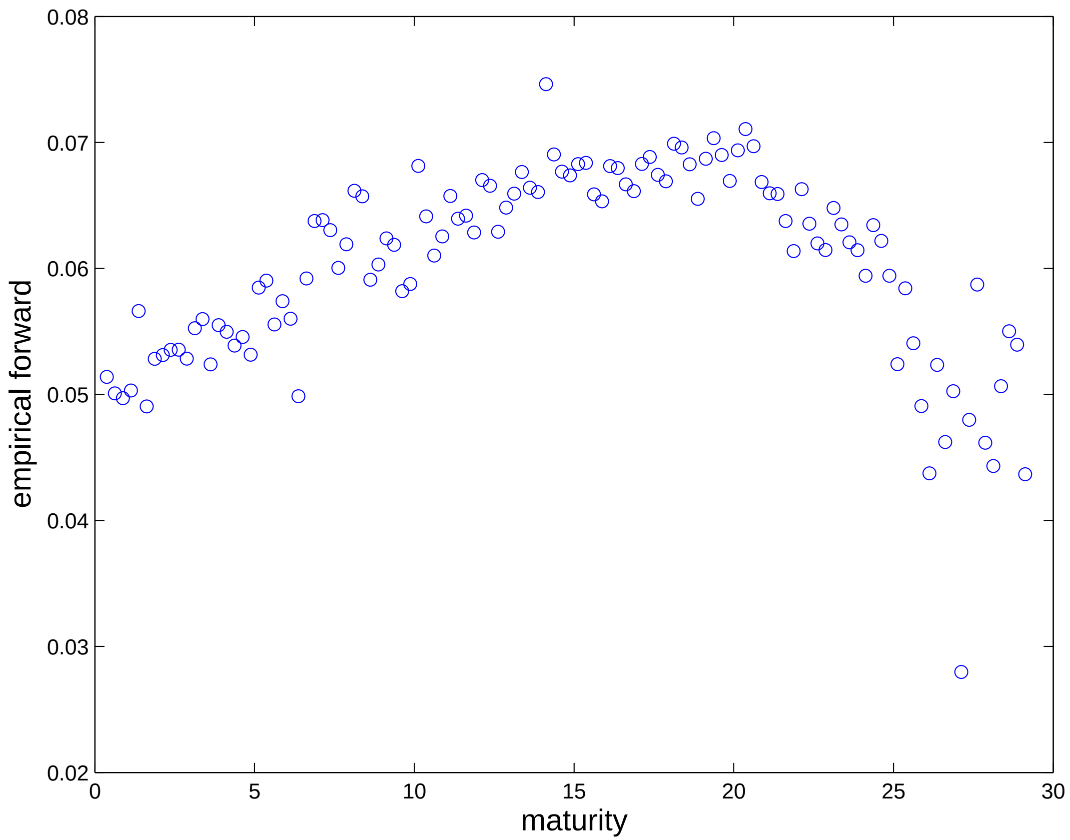
$$D(t) = \exp\left\{ -\int_0^t f(s)ds \right\} \text{ for all } t$$

$$f(t) = -\frac{d}{dt}\log\{D(t)\}$$

$$\text{empirical forward } = -\frac{\log\{P(t_{i+1})\} - \log\{P(t_i)\}}{t_{i+1} - t_i}$$

$$P(t) = \text{observed price at time } t$$

Empirical Forward Rate



STRIPS on Dec 31, 1995: empirical forward rate

Modelling Coupon Bonds

- $P_1, \cdots, P_n$ denote observed market prices of n bonds (coupon or zero-coupon)
- Bond i has fixed payments $C_i(t_{i,j})$ due on dates $t_{i,j}, j = 1, \ldots, N_i$ ($N_i = 1$ for zero-coupon bonds)
- Model price for the ith coupon bond:

$$\widehat{P}_i(\boldsymbol{\delta}) = \sum_{j=1}^{N_i} C_i(t_{i,j}) \exp\left\{ - \int_0^{t_{i,j}} f(s, \boldsymbol{\delta}) ds \right\}$$

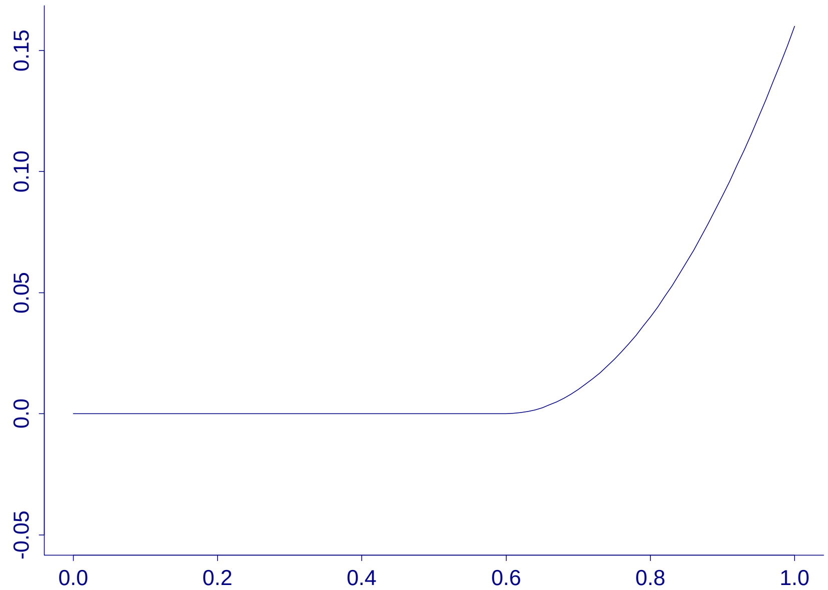
$f(\cdot, \boldsymbol{\delta})$ is a model for the forward rate

Spline Model of Forward Rate

- $f(s, \boldsymbol{\delta}) = \boldsymbol{\delta}^{\mathsf{T}} \mathbf{B}(s)$
  - $\mathbf{B}(s)$ is a vector of spline basis functions
  - $\boldsymbol{\delta}$ is a vector of spline coefficients
- $\therefore\ F(t, \boldsymbol{\delta}) := \int_0^t f(s, \boldsymbol{\delta}) ds = ty(t, \boldsymbol{\delta}) = \boldsymbol{\delta}^{\mathsf{T}} \mathbf{B}^I(s)$
  - $\mathbf{B}^I(t) := \int_0^t \mathbf{B}(s) ds$.

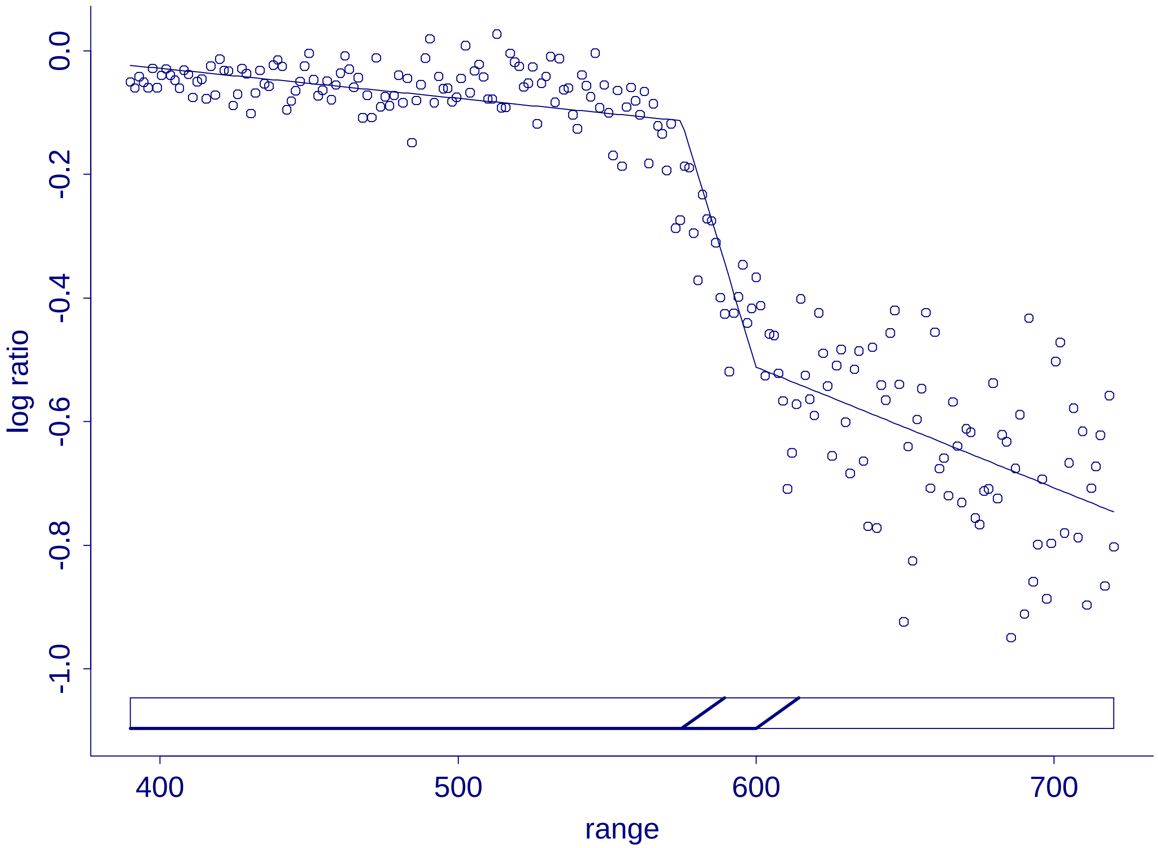
Example: Quadratic Splines

$$\mathbf{B}(s) = \left(1,\ s,\ s^2,\ (s-\kappa_1)_+^2, \ldots, (s-\kappa_K)_+^2\right)^{\mathsf{T}}$$

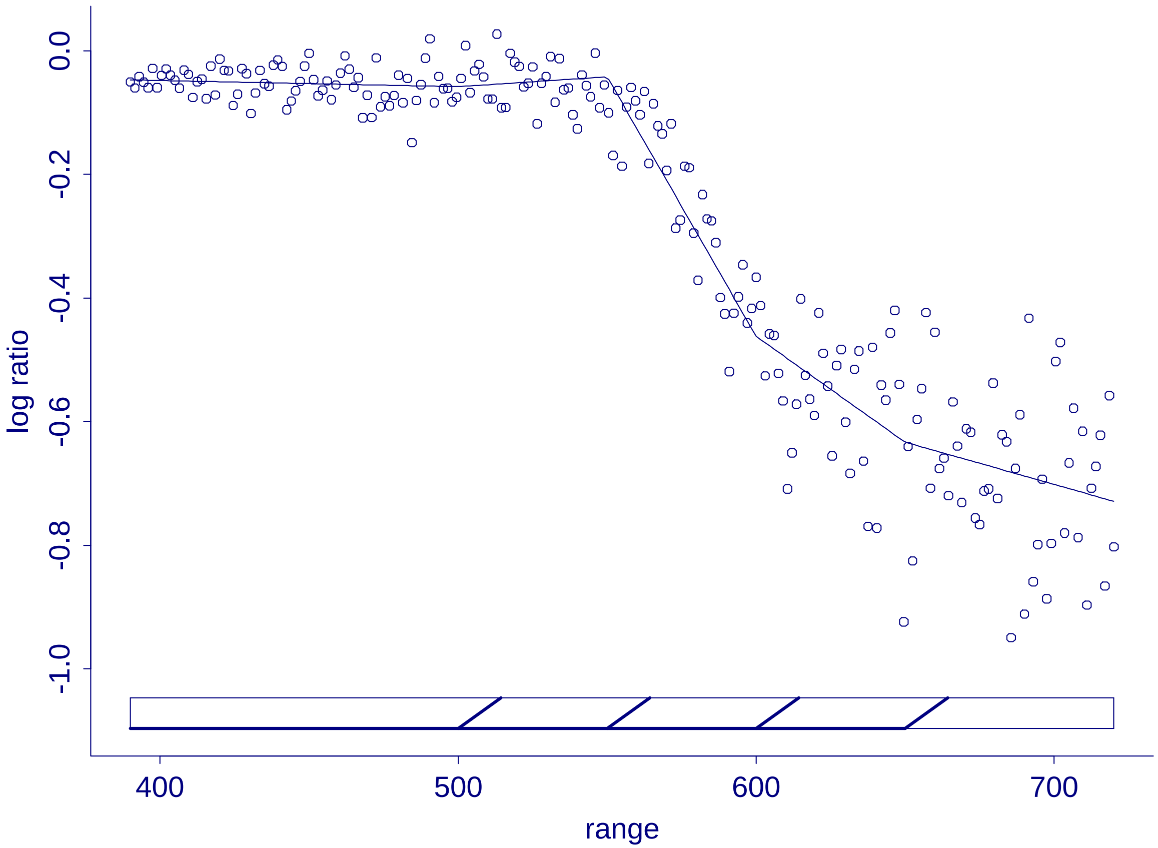


Plus function with knot at 0.6

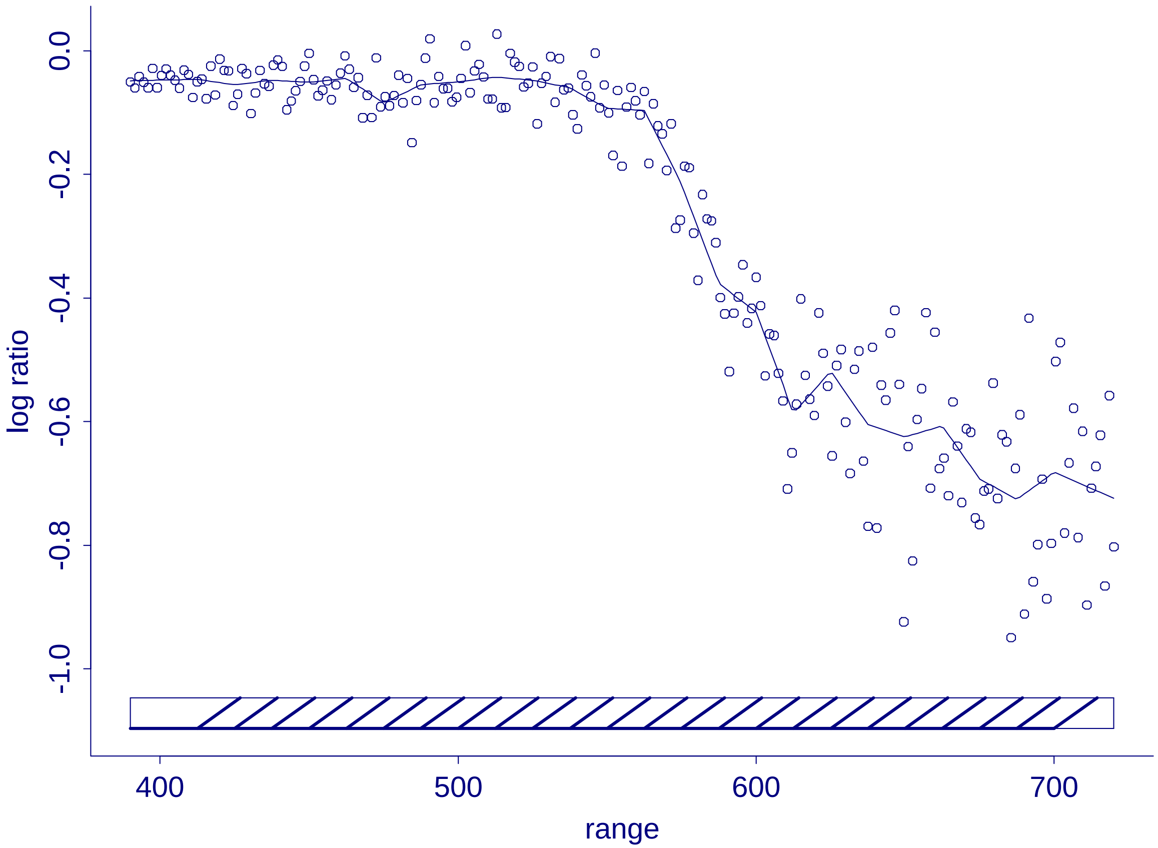
Linear Spline – 2 Knots



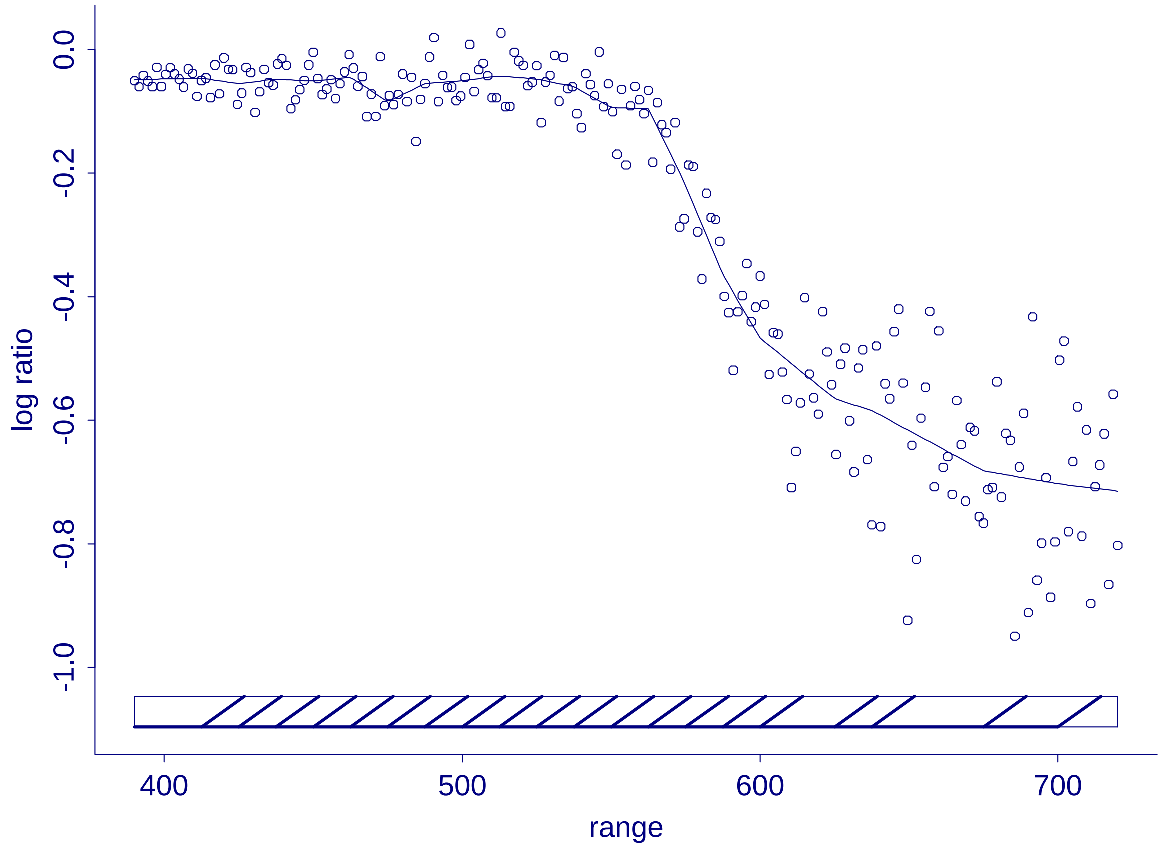
Linear Spline – 4 Knots



Linear Spline – 24 Knots



Lidar Data — Carefully Chosen Knots



There *is* a better way to get a smooth fit than selecting knots

Modelling the Forward Rate

From before:

$$\mathbf{B}(t) = \left(\ 1,\ t,\ \cdots,\ t^p,\ (t-\kappa_1)_+^p,\ \cdots,\ (t-\kappa_K)_+^p \ \right)^{\mathsf{T}}$$

Therefore:

$$\mathbf{B}^I(t) := \int_0^t \mathbf{B}(s)ds = \left(t \quad \ldots \quad \tfrac{t^{p+1}}{p+1} \quad \tfrac{(t-\kappa_1)_+^{p+1}}{p+1} \quad \ldots \quad \tfrac{(t-\kappa_K)_+^{p+1}}{p+1} \right)^{\mathsf{T}}.$$

Penalized Least-Squares

$$Q_{n,\lambda}(\boldsymbol{\delta}) = \frac{1}{n}\sum_{i=1}^{n}\left[h(P_i) - h\{\widehat{P}_i(\boldsymbol{\delta})\}\right]^2 + \lambda\boldsymbol{\delta}^{\mathsf{T}}\mathbf{G}\boldsymbol{\delta}$$

or equivalently

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

- h is a monotonic transformation: "transform-both-sides" model
- $\lambda\boldsymbol{\delta}^{\mathsf{T}}\mathbf{G}\boldsymbol{\delta}$ is a "roughness" penalty
  - $\lambda \geq 0$
  - $\mathbf{G}$ is positive semi-definite

Penalized Least-Squares

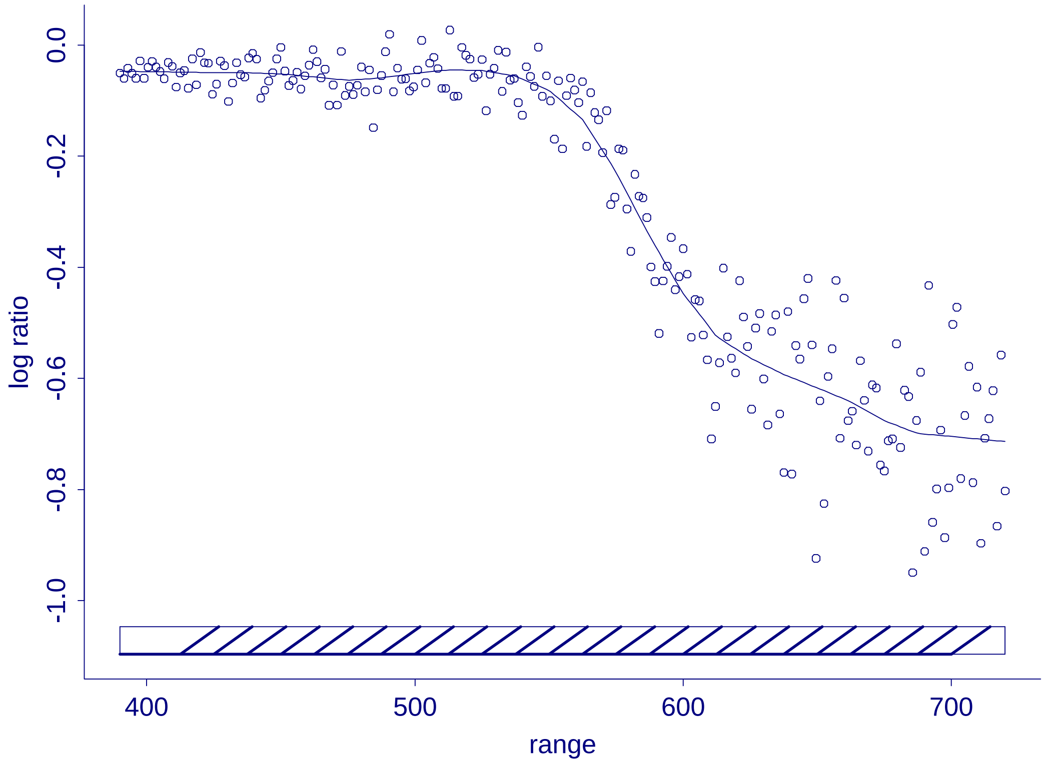
From previous slide:

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

Several sensible choices for $\mathbf{G}$

1. $\mathbf{G}$ is a diagonal matrix
   - last K diagonal elements equal to one
   - all others zero.
   - penalizes jumps at the knots in the pth derivative of the spline.
2. quadratic penalty on the dth derivative $\int\{f^{(d)}(s)\}^2\,ds$
   - uses $G_{ij} = \int B_j^{(d)}(t)B_k^{(d)}(t)dt$
     - $B_j(t)$ is the jth element of $\mathbf{B}(t)$

Linear Spline with 24 Knots Fit by Penalized Least Squares



- Number of knots has little effect on fit provide it is at least 15
- Choice of λ is crucial

Using Zero Coupon Bonds

- Now assume we are using zeros, e.g., STRIPS
- P_i has a single payment of \$1 at time t_i
- Therefore,

$$Q_{n,\lambda}(\boldsymbol{\delta}) = \frac{1}{n}\sum_{i=1}^{n}\left(h(P_i) - h\left[\exp\left\{-\boldsymbol{\delta}^{\mathsf{T}}\mathbf{B}^{I}(t_i)\right\}\right]\right)^2 + \lambda\boldsymbol{\delta}^{\mathsf{T}}\mathbf{G}\boldsymbol{\delta}$$

Choosing the Knots

- κ_k is the $\frac{k}{(K+1)}$th sample quantile of $\{t_i\}_{i=1}^n$
- the t_i are nearly equally spaced so the knots are also

Effective Number of Parameters of a Fit

There exists a matrix $\mathbf{S}(\lambda)$ such that

$$\begin{pmatrix} \widehat{P}_1 \\ \vdots \\ \widehat{P}_n \end{pmatrix} \approx \mathbf{S}(\lambda) \begin{pmatrix} P_1 \\ \vdots \\ P_n \end{pmatrix}$$

- $\mathbf{S}(\lambda)$ is called the smoother matrix or hat matrix
- $\mathsf{DF}(\lambda) := \mathrm{trace}\{\mathbf{S}(\lambda)\}$ is called the degrees of freedom of the fit or the effective number of parameters

Generalized Cross-Validation

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

- one chooses λ to minimize GCV(λ)
- θ is a user-specified tuning parameter
- $\theta = 1$ is ordinary GCV
- Fisher, Nychka, and Zervos used $\theta = 2$
  - this causes more smoothing
  - Question: why doesn't ordinary GCV work well here?

EBBS

- To estimate MSE add together:
  - estimated squared bias
  - estimated variance
- Gives $\text{MSE}(\widehat{f};\, t, \lambda)$, the estimated MSE of $\widehat{f}$ at t and λ.
  - then $\sum_{i=1}^{n} \text{MSE}(\widehat{f};\, t_i, \lambda)$ is minimized over λ
- EBBS estimates bias at any fixed t by
  - computing the fit at t for a range of values of the smoothing parameter
  - fitting a curve to model bias

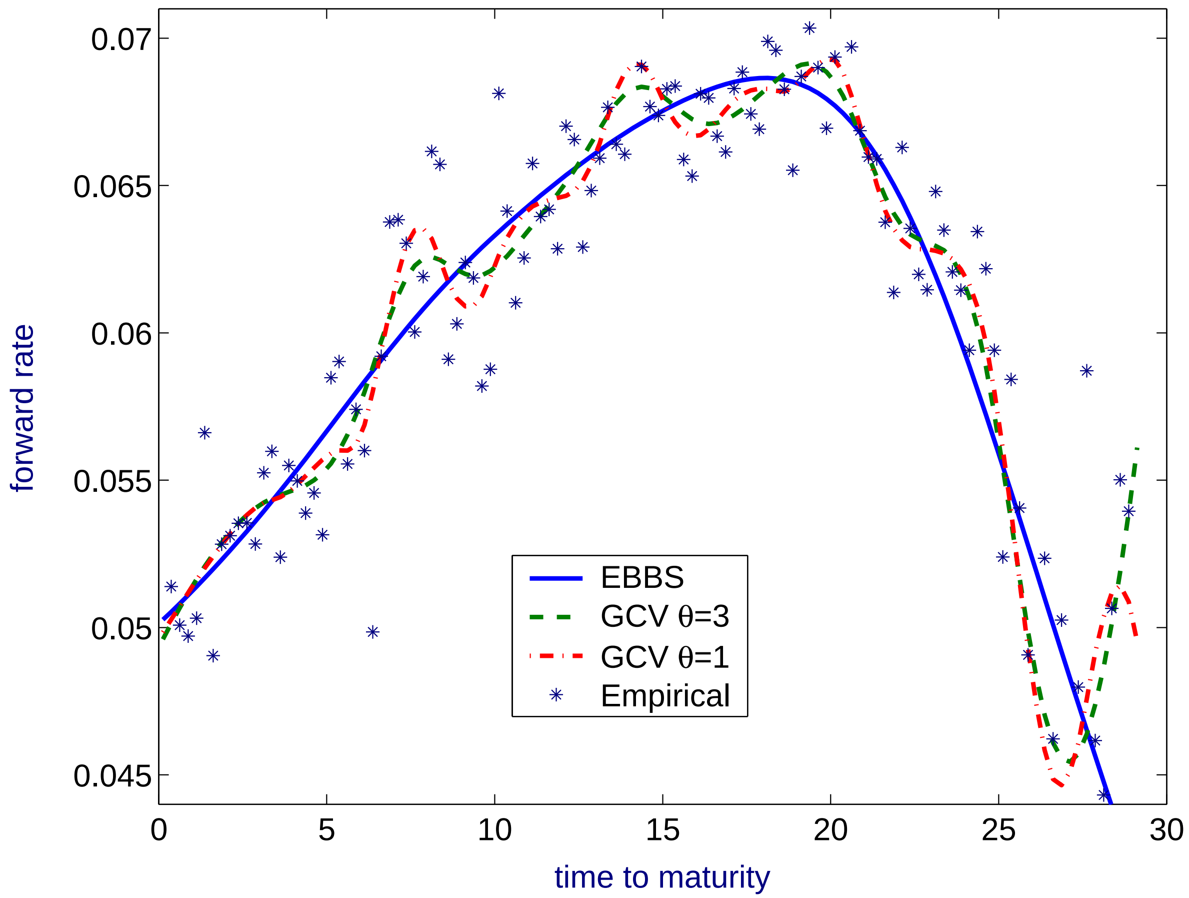
EBBS – Estimating Bias

- to the first order, the bias is $\gamma(t)\lambda$ for some $\gamma(t)$
- Let $\widehat{f}(t, \lambda)$ be $\widehat{f}$ depending on maturity and λ
- Compute $\left\{\lambda_\ell, \widehat{f}(t, \lambda_\ell)\right\}, \ \ell = 1, \ldots, L$
  - $\lambda_1 < \ldots < \lambda_L$ is the grid of values of λ
  - we used $L = 50$ values of λ
  - $\log_{10}(\lambda_\ell)$ were equally spaced between -7 and 1
  - DF$(10) = 4.8$ and DF$(10^{-7}) = 28.9$ for a 40-knot cubic spline fit

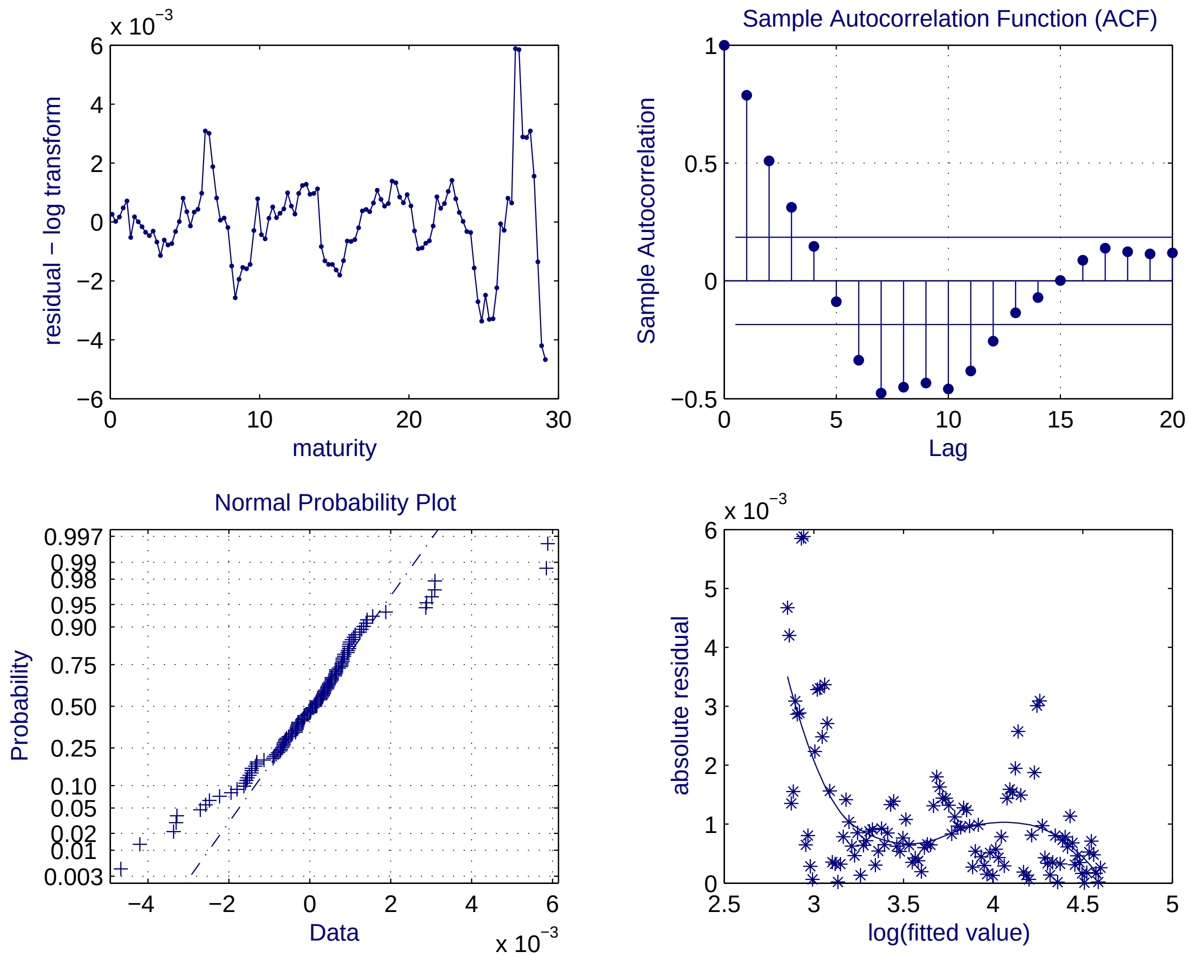
EBBS – Estimating Bias

- For any fixed t, fit a straight line to the data $\{(\lambda_i, \widehat{f}(t, \lambda_i) : i = 1, \ldots, L\}$
- slope of the line is $\widehat{\gamma}(t)$
- estimate of squared bias at t and λ_ℓ is $(\widehat{\gamma}(t)\,\lambda_\ell)^2$

EBBS versus GCV

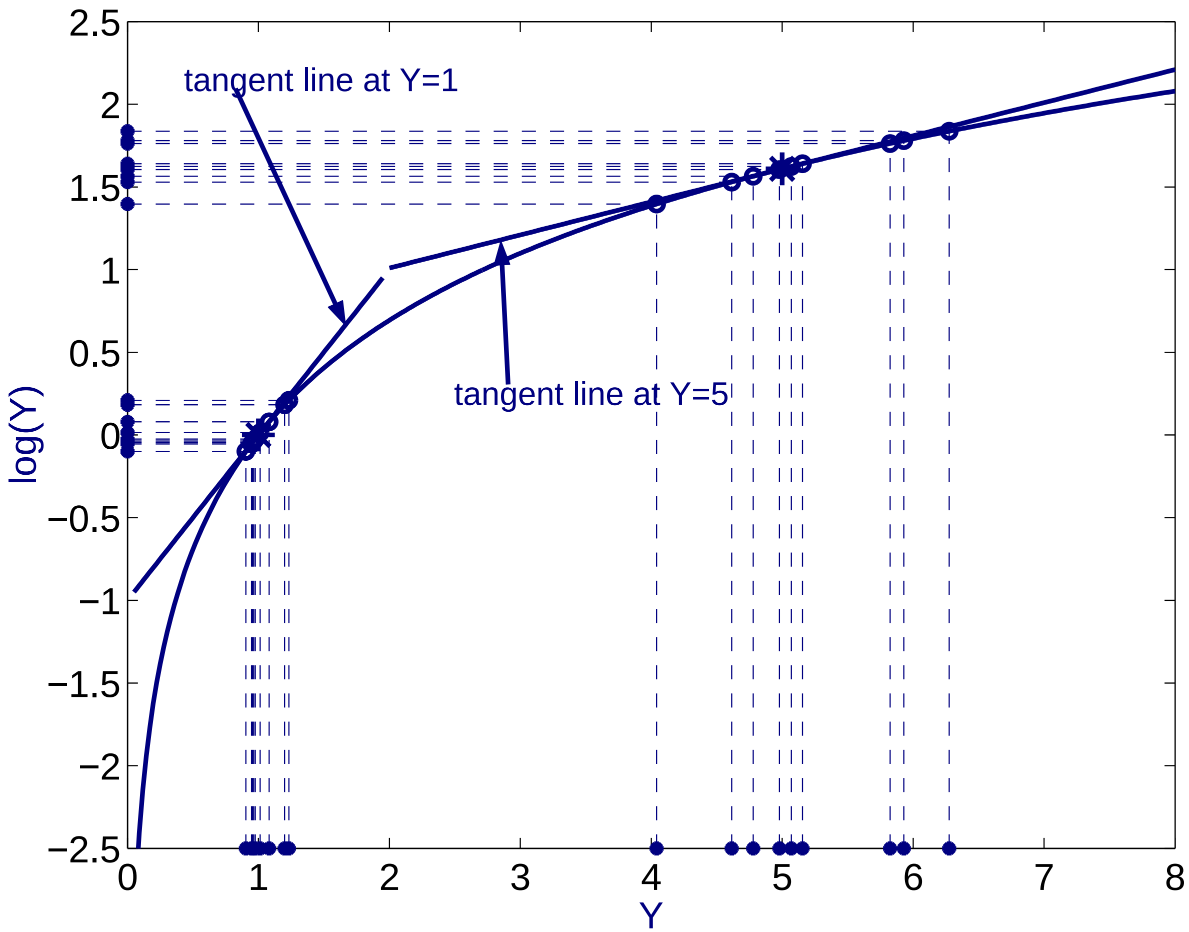


EBBS Fit: Residual Analysis



Residuals from fit using $h(\cdot) = \log(\cdot)$

Geometry of Transformations



Strength of a Transformation

- Suppose $y_1 < y_2$
- strength of a transformation h:

$$\text{strength} = \frac{h'(y_2)}{h'(y_1)} - 1$$

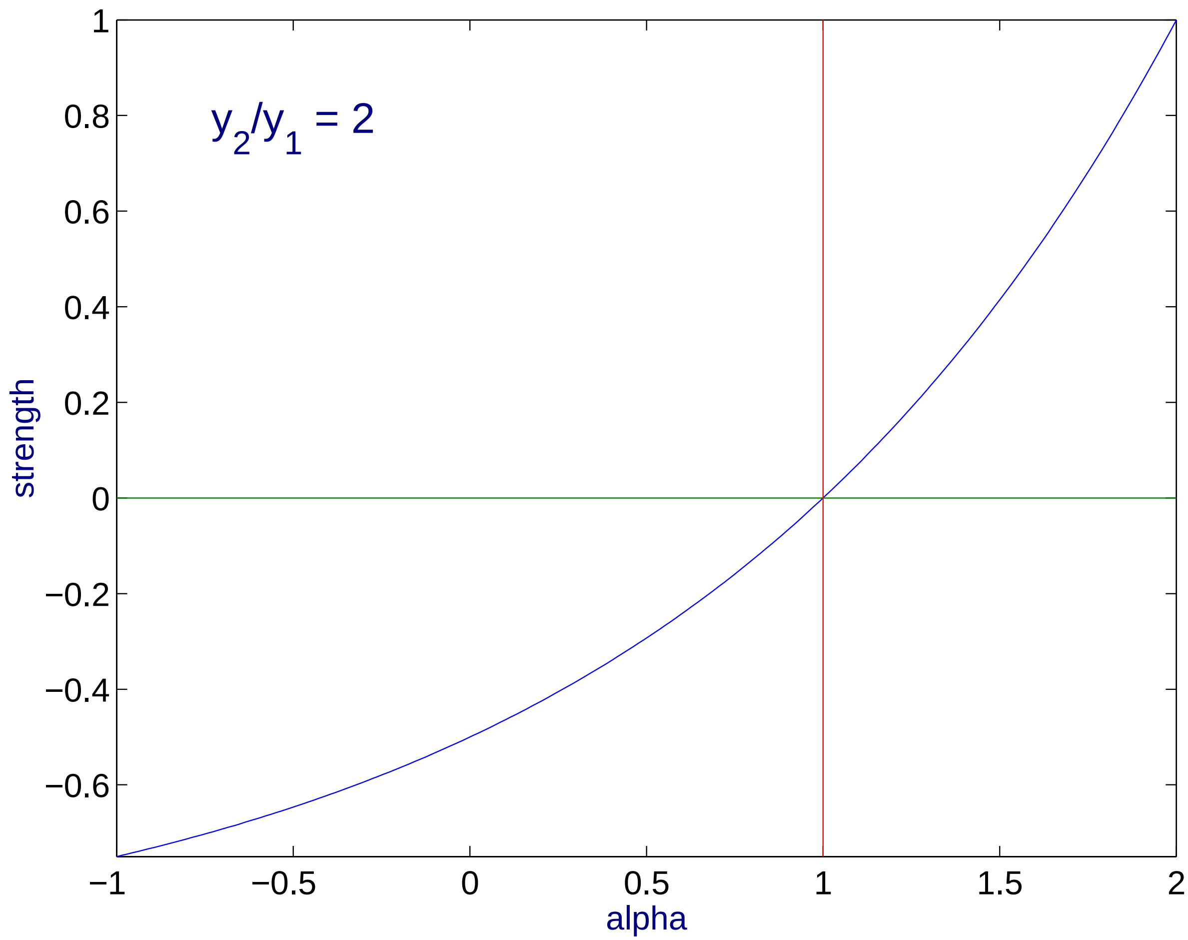
- Example:

$$\begin{aligned} h(y;\alpha) &= \frac{y^\alpha - 1}{\alpha} \text{ if } \alpha \neq 0 \\ &= \log(y) \text{ if } \alpha = 0 \end{aligned}$$

-

$$\begin{aligned} \text{strength} &:= \left(\frac{y_2}{y_1}\right)^{\alpha-1} - 1 \\ &> 0 \text{ if } \alpha > 1 \\ &< 0 \text{ if } \alpha < 1 \end{aligned}$$

Strength of a Transformation



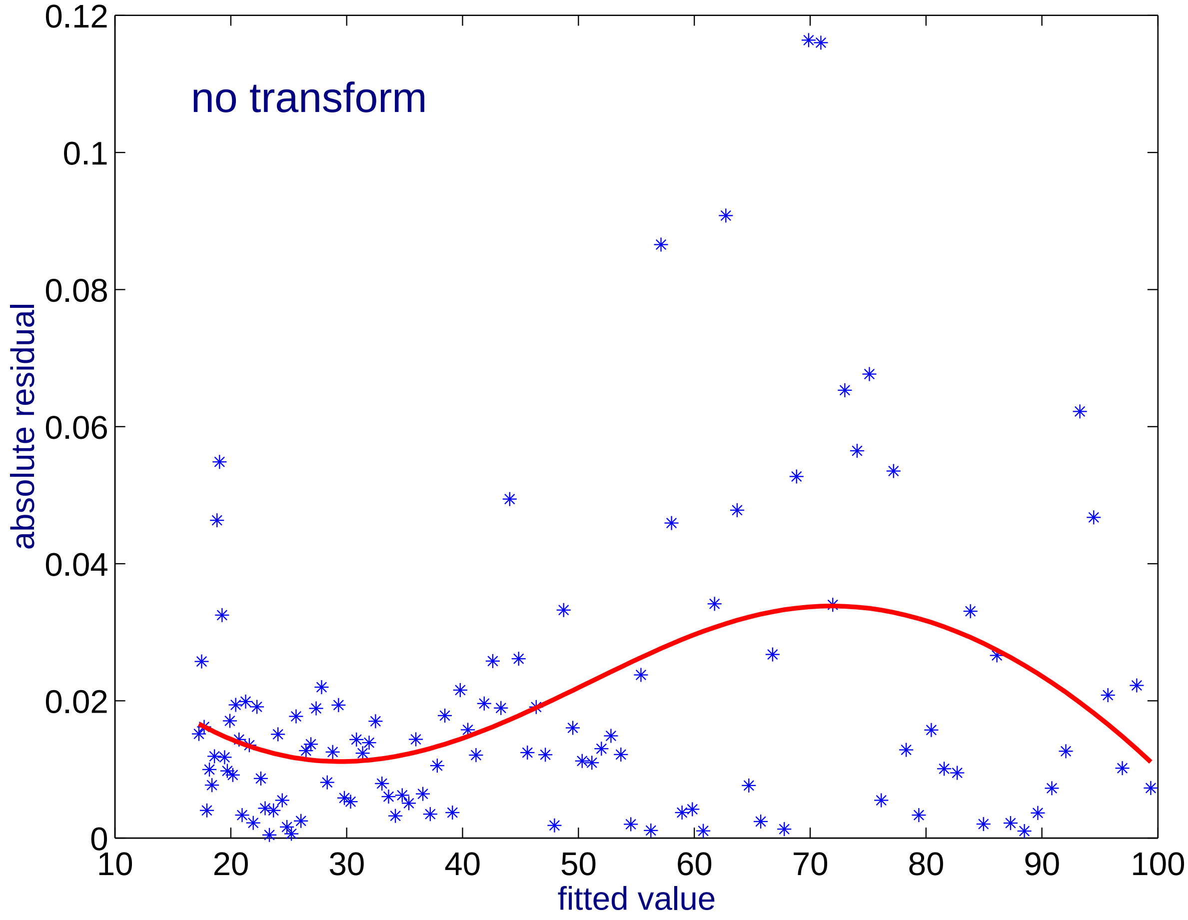
Transformation and Weighting

- log is the linearizing transformation
  - convenient
  - induces some heteroscedasticity, but not enough to cause a problem
- $\log\{P(t)\}/t = -$ yield
  - cause severe heteroscedasticity – avoid

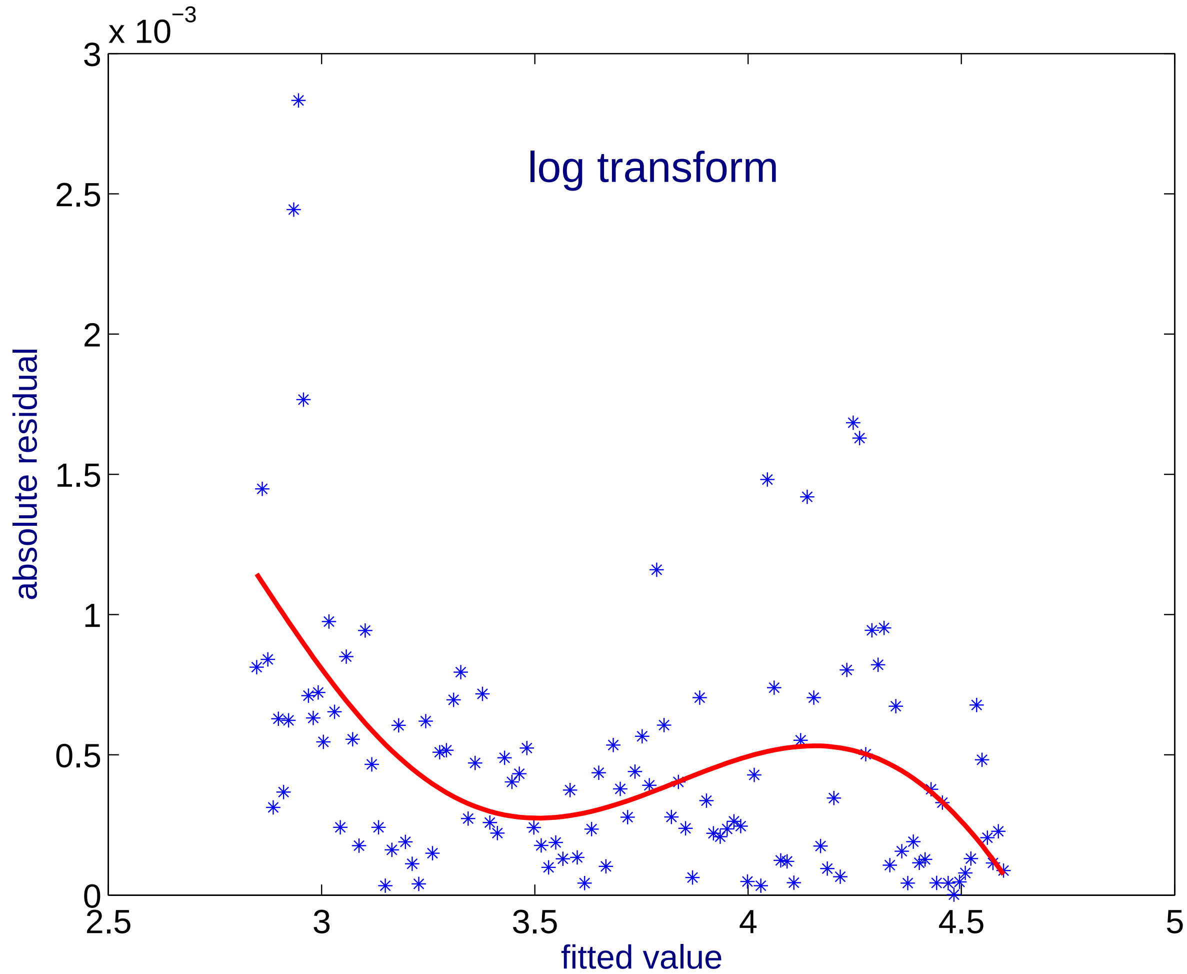
Transformation and weighting should be done primarily **to induce the assumed noise distribution**, which is:

- normal
- constant variance

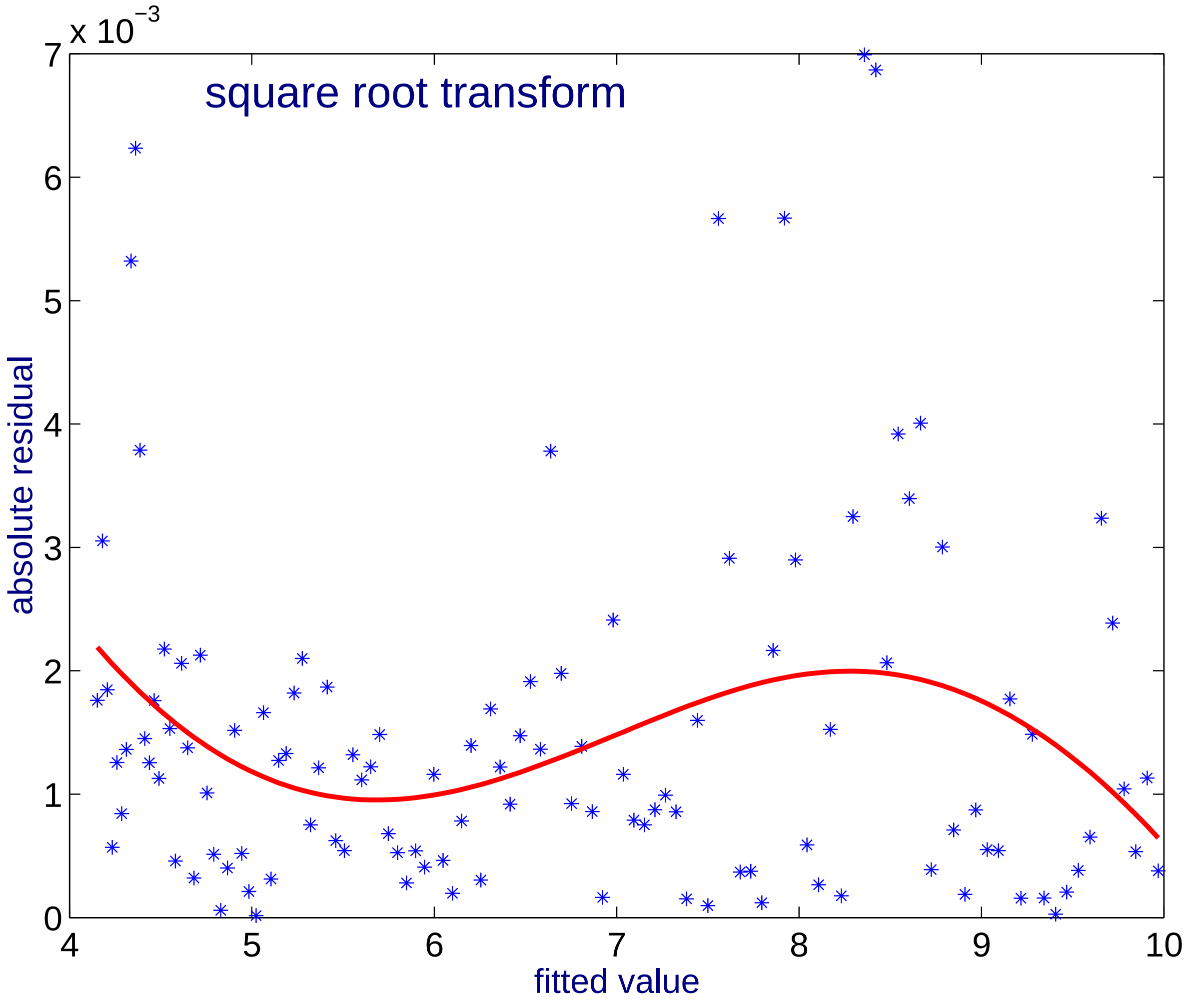
Transformation and Weighting



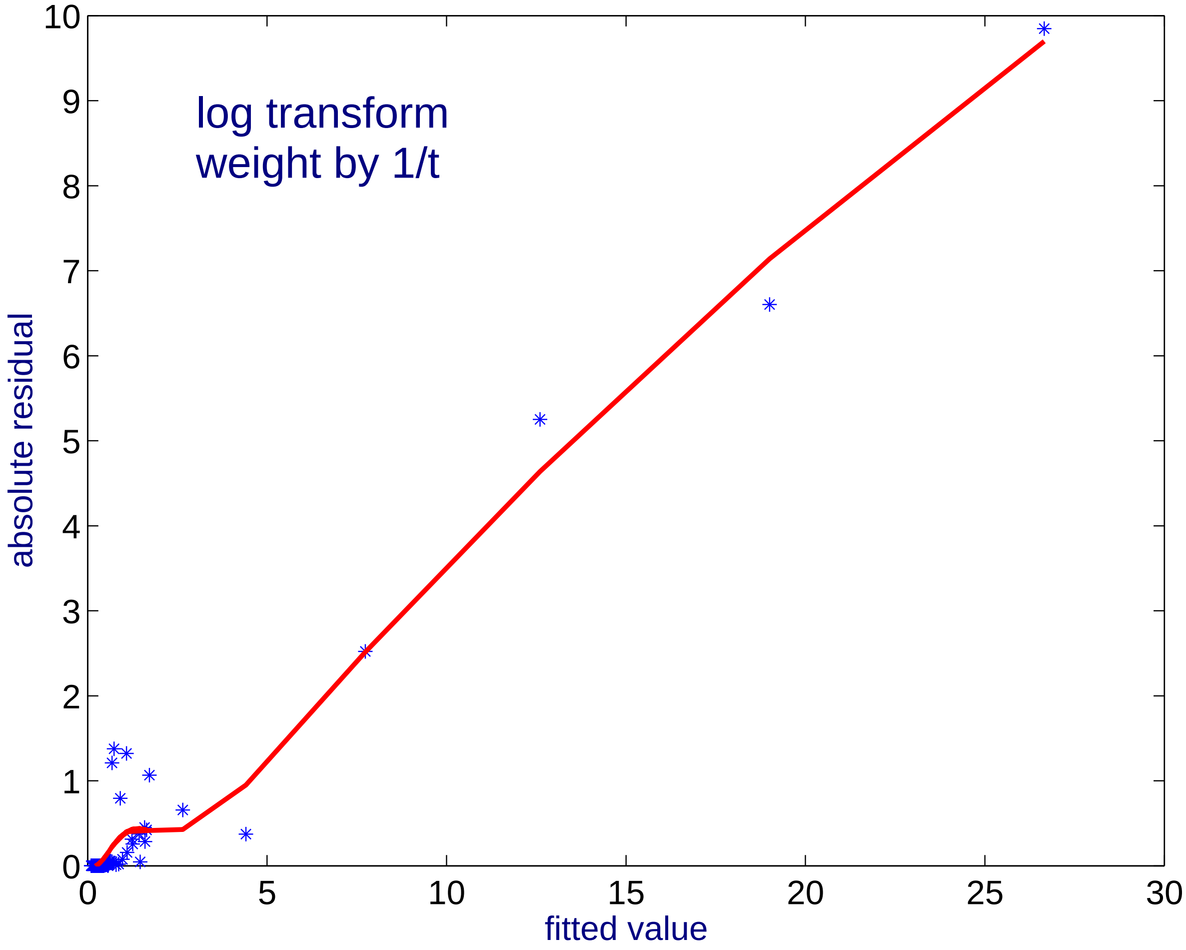
Transformation and Weighting



Transformation and Weighting



Transformation and Weighting



Modelling the Correlation

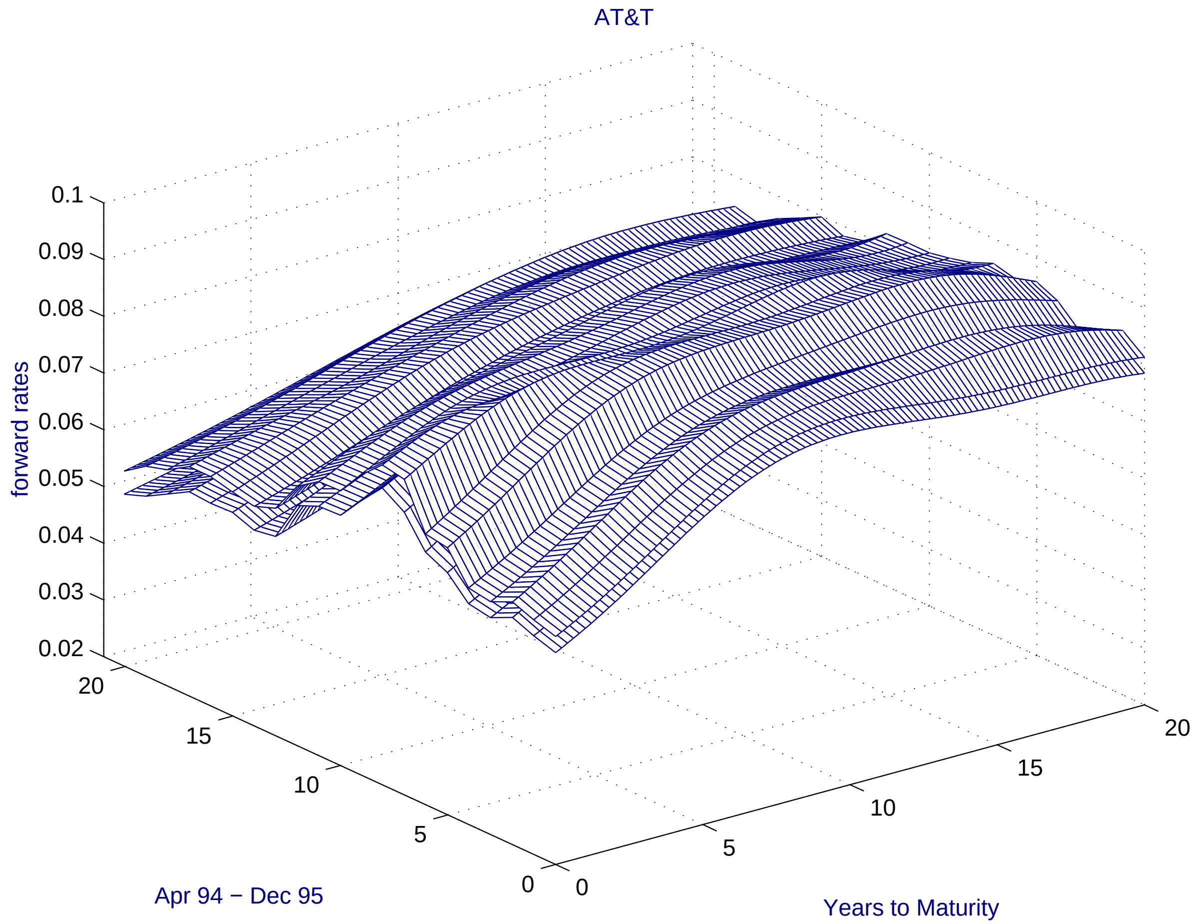
- open problem
- probably not stationary
- simulations show that stationary AR and MA processes do not have the same problem with GCV as seen with actual price data

Modelling Corporate Term Structure

$$f_C(t) = f_{Tr}(t) + \alpha_0 + \alpha_1 t + \alpha_2 t^2$$

- $\alpha_0 + \alpha_1 t + \alpha_2 t^2$ is the credit spread
- H_0: $\alpha_1 = \alpha_2 = 0$ is accepted for AT&T data
- $\alpha_0 > 0$ for the AT&T data

Modelling Corporate Term Structure



Modelling Corporate Term Structure

Question: Should one smooth over both date and time to maturity?

Asymptotics

The PLS estimator is the solution to

$$\sum_{i=1}^{n} \psi_i(\boldsymbol{\delta}, \lambda, \mathbf{G}) = 0$$

for an appropriate $\psi_i(\cdot, \cdot, \cdot)$

Asymptotics: $\lambda \to 0$

Theorem 1

- *let* $\{\widehat{\boldsymbol{\delta}}_{n,\lambda_n}\}$ *be a sequence of penalized least squares estimators*
- *assume typical "regularity" assumptions*
- *suppose* λ_n *is* $o(1)$
- *then* $\widehat{\boldsymbol{\delta}}_n$ *is a (strongly) consistent for* $\boldsymbol{\delta}_0$
- *if* λ_n *is* $o(n^{-1/2})$*, then*

$$\sqrt{n}\left(\widehat{\boldsymbol{\delta}}_{n,\lambda_n} - \boldsymbol{\delta}_0\right) \xrightarrow{D} N\left\{0, \sigma^2\Omega^{-1}(\boldsymbol{\delta}_0)\right\},$$

where

$$\Omega(\boldsymbol{\delta}_0) := \lim_n \boldsymbol{\Sigma}_n, \;\; \boldsymbol{\Sigma}_n = \sigma^{-2}n^{-1}\sum_{i=1}^{n} E\left\{\psi_i(\boldsymbol{\delta}, \lambda, \mathbf{G})\psi_i(\boldsymbol{\delta}, \lambda, \mathbf{G})^{\mathsf{T}}\right\}$$

Asymptotics: λ fixed

- assume $\lambda_n \equiv \lambda$
- the bias does not shrink to 0
  - limit of $\widehat{\boldsymbol{\delta}}_{n,\lambda}$ solves

$$\lim_{n\to\infty} E\left\{ n^{-1} \sum_{i=1}^{n} \psi_i(\boldsymbol{\delta}, \lambda, \mathbf{G}) \right\} = 0$$

- the large sample variance formula is

$$\widehat{\mathrm{Var}}\{\widehat{\boldsymbol{\delta}}(\lambda)\} = \frac{\sigma^2}{n} \left[\{\boldsymbol{\Sigma}_n + \lambda \mathbf{G}\}^{-1} \boldsymbol{\Sigma}_n \{\boldsymbol{\Sigma}_n + \lambda \mathbf{G}\}^{-1} \right].$$

Summary

- splines are convenient for estimating term structure
- penalization is better, or at least easier, than knot selection
- EBBS provides a reasonable amount of smoothing
- GCV undersmooths because
  - noise is correlated
  - target function is a derivative
- corporate term structure can be estimated by "borrowing strength" from treasury bonds
- a constant credit spread fits the data reasonably well
- asymptotics are available for inference

References

Jarrow, R., Ruppert, D., and Yu, Y. (2004) Estimating the interest rate term structure of corporate debt with a semiparametric penalized spline model, *JASA*, to appear.

Available at:

`http://www.orie.cornell.edu/~davidr`

- see "Recent Papers"
- also see "Recent Talks" for these slides

References

Ruppert, D. (2004) *Statistics and Finance: An Introduction*, Springer, New York – splines, term structure, transformations

Ruppert, D., Wand, M.P., and Carroll, R.J. (2003) *Semiparametric Regression*, Cambridge University Press, New York – splines

Carroll, R.J. and Ruppert, D. (1988), *Transformation and Weighting in Regression*, Chapman & Hall, New York. – transform-both-sides