

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

Slide 1

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 .

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

Slide 3

- $f(t)$ 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.,

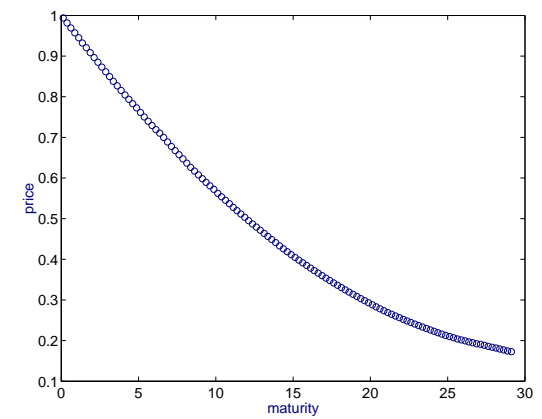
$$y(t) = \frac{1}{t} \int_0^t f(s) ds = -\frac{1}{t} \log\{D(t)\}$$

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

Slide 2

Discount Function, Forward Rates, and Yields

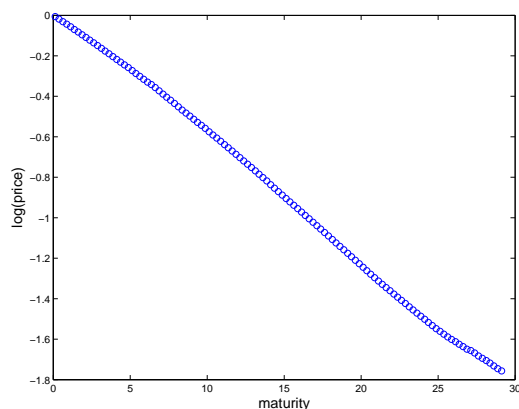


Slide 4

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

Prices, Forward Rates, and Yields

Slide 5



STRIPS on Dec 31, 1995: log prices

Empirical Forward Rate

Slide 7

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

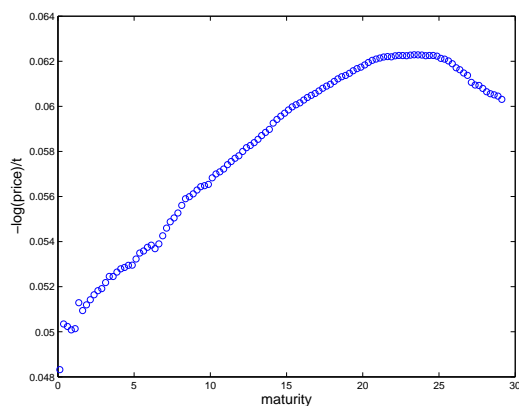
$$f(t) = - \frac{d}{dt} \log \{ D(t) \}$$

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

$P(t)$ = observed price at time t

Discount Function, Forward Rates, and Yields

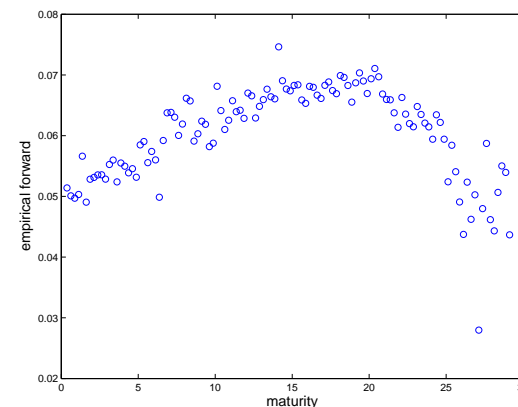
Slide 6



STRIPS on Dec 31, 1995: empirical yields

Empirical Forward Rate

Slide 8



STRIPS on Dec 31, 1995: empirical forward rate

- P_1, \dots, 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 $t_{i,j}, j = 1, \dots, N_i$ ($N_i = 1$ for zero-coupon bonds)
- Model price for the i th coupon bond:

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

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

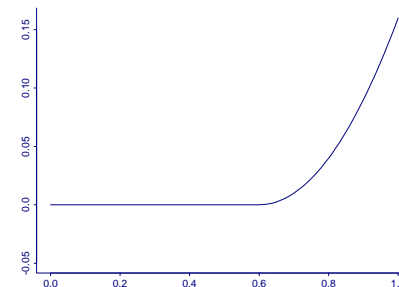
Slide 9

Spline Model of Forward Rate

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

Slide 10

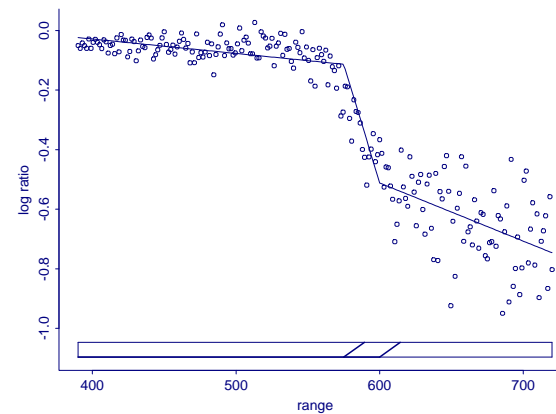
$$\mathbf{B}(s) = (1, s, s^2, (s - \kappa_1)_+^2, \dots, (s - \kappa_K)_+^2)^T$$



Plus function with knot at 0.6

Slide 11

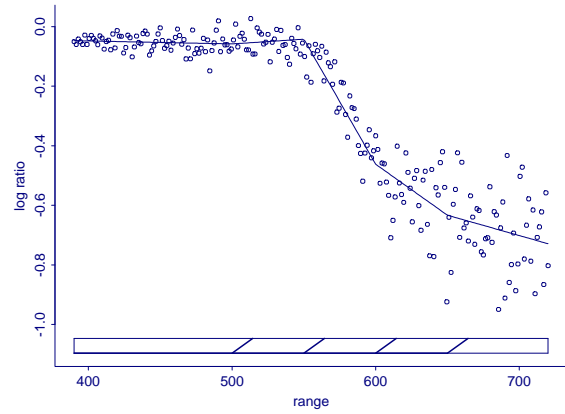
Linear Spline – 2 Knots



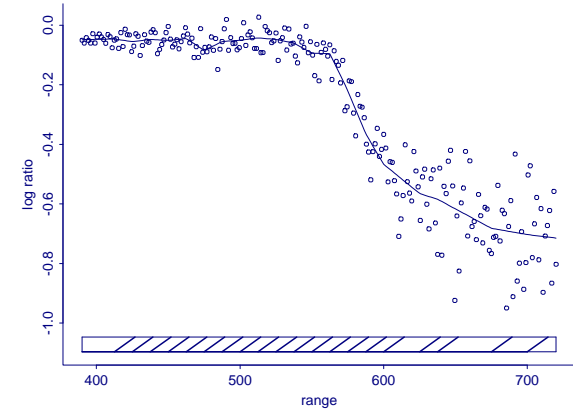
Slide 12

Linear Spline – 4 Knots

Slide 13



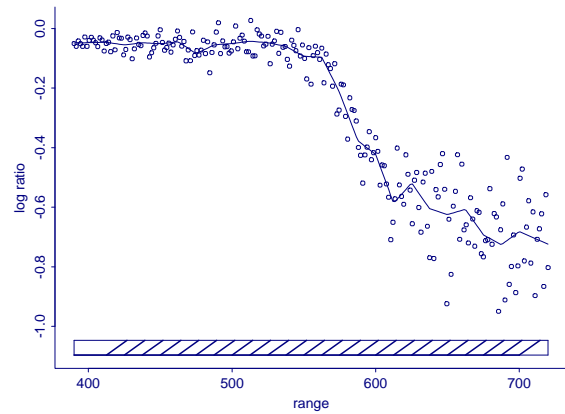
Slide 15



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

Linear Spline – 24 Knots

Slide 14



Slide 16

Modelling the Forward Rate

From before:

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

Therefore:

$$\mathbf{B}^I(t) := \int_0^t \mathbf{B}(s) ds = \left(t \quad \dots \quad \frac{t^{p+1}}{p+1} \quad \frac{(t - \kappa_1)_+^{p+1}}{p+1} \quad \dots \quad \frac{(t - \kappa_K)_+^{p+1}}{p+1} \right)^\top.$$

Penalized Least-Squares

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

or equivalently

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

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

Slide 17

Penalized Least-Squares

From previous slide:

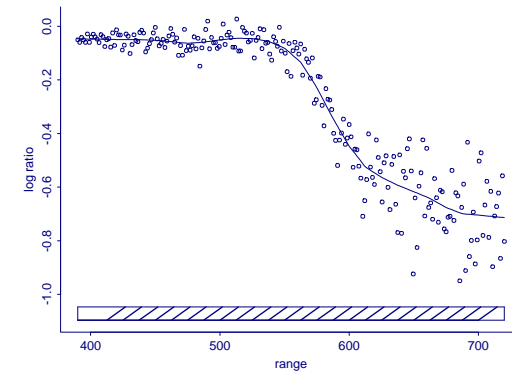
$$Q_{n,\lambda}(\boldsymbol{\delta}) = \frac{1}{n} \sum_{i=1}^n \left\{ h(P_i) - h \left[\sum_{j=1}^{N_i} C_i(t_{i,j}) \exp \left\{ -\boldsymbol{\delta}^T \mathbf{B}^I(t_{i,j}) \right\} \right] \right\}^2 + \lambda \boldsymbol{\delta}^T \mathbf{G} \boldsymbol{\delta}.$$

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 p th derivative of the spline.
2. quadratic penalty on the d th 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 j th element of $\mathbf{B}(t)$

Slide 18

Linear Spline with 24 Knots Fit by Penalized Least Squares



Slide 19

- 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,

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

Slide 20

Slide 21

- 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

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

Slide 23

- one chooses λ to minimize $GCV(\lambda)$
- 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?

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

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

Slide 22

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

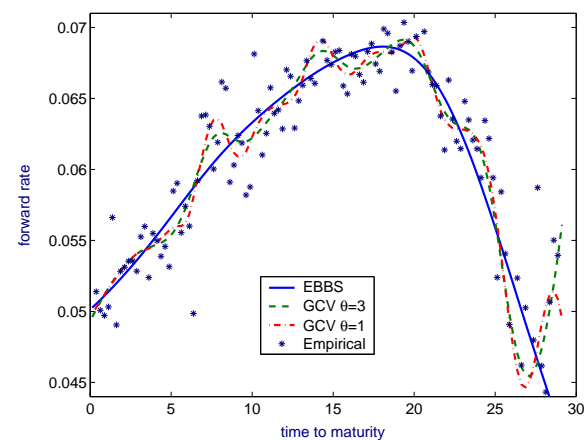
Slide 24

- To estimate MSE add together:
 - estimated squared bias
 - estimated variance
- Gives $\text{MSE}(\hat{f}; t, \lambda)$, the estimated MSE of \hat{f} at t and λ .
 - then $\sum_{i=1}^n \text{MSE}(\hat{f}; t_i, \lambda)$ 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

Slide 25

- to the first order, the bias is $\gamma(t)\lambda$ for some $\gamma(t)$
- Let $\hat{f}(t, \lambda)$ be \hat{f} depending on maturity and λ
- Compute $\{\lambda_\ell, \hat{f}(t, \lambda_\ell)\}, \ell = 1, \dots, L$
 - $\lambda_1 < \dots < \lambda_L$ 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

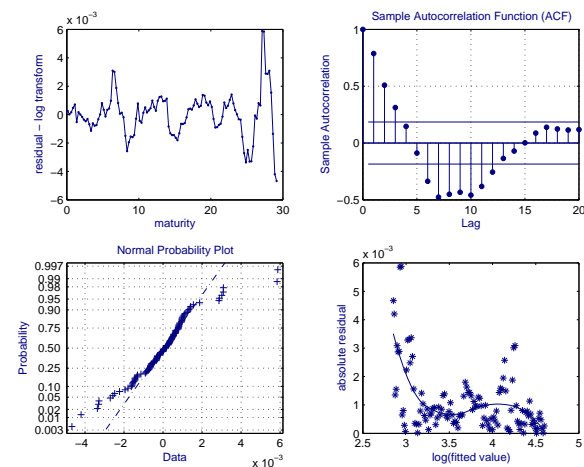
Slide 27



Slide 26

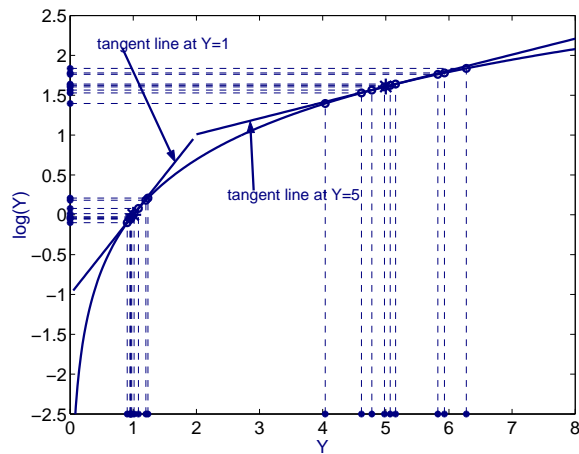
- For any fixed t , fit a straight line to the data $\{(\lambda_i, \hat{f}(t, \lambda_i) : i = 1, \dots, L)\}$
- slope of the line is $\hat{\gamma}(t)$
- estimate of squared bias at t and λ_ℓ is $(\hat{\gamma}(t) \lambda_\ell)^2$

Slide 28

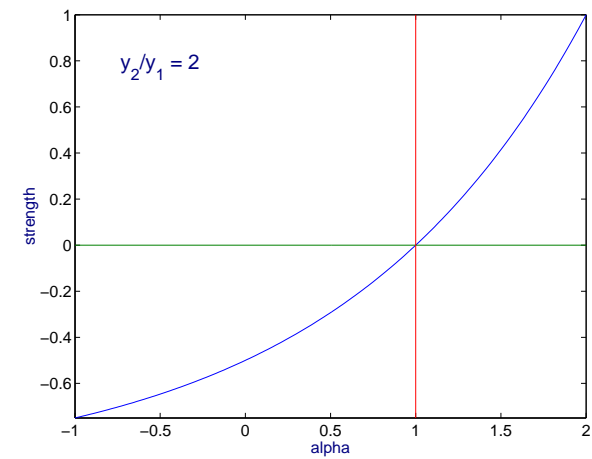


Residuals from fit using $h(\cdot) = \log(\cdot)$

Slide 29



Slide 31



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:

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

-

$$\text{strength} := \left(\frac{y_2}{y_1}\right)^{\alpha-1} - 1$$

$$\begin{aligned} &> 0 \text{ if } \alpha > 1 \\ &< 0 \text{ if } \alpha < 1 \end{aligned}$$

Slide 30

Transformation and Weighting

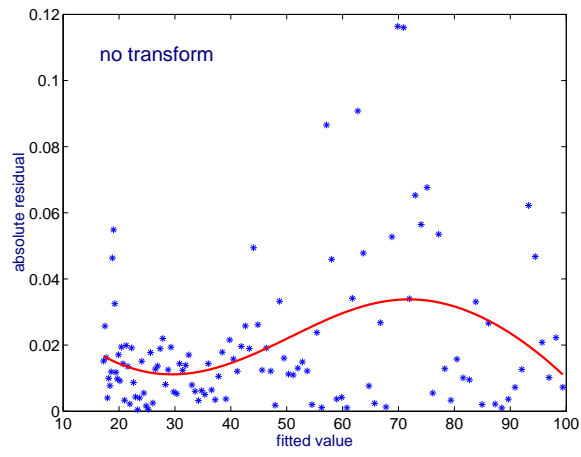
- log is the linearizing transformation
 - convenient
 - induces some heteroscedasticity, but not enough to cause a problem
- $\log\{P(t)\}/t = - \text{yield}$
 - cause severe heteroscedasticity – avoid

Slide 32

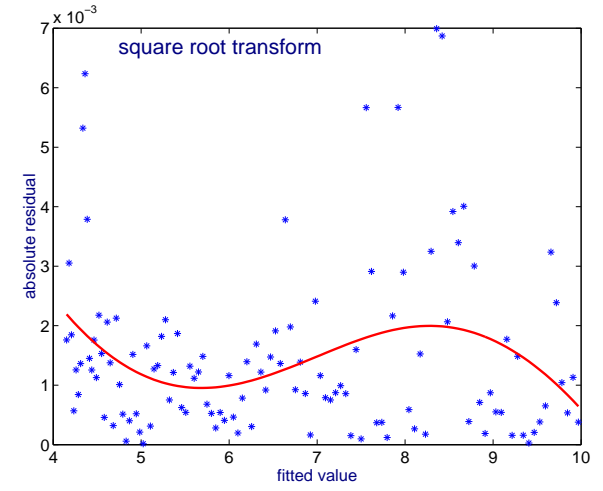
Transformation and weighting should be done primarily to induce the assumed noise distribution, which is:

- normal
- constant variance

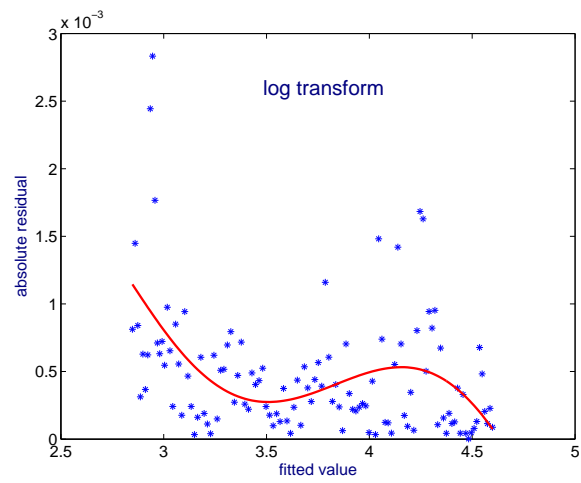
Slide 33



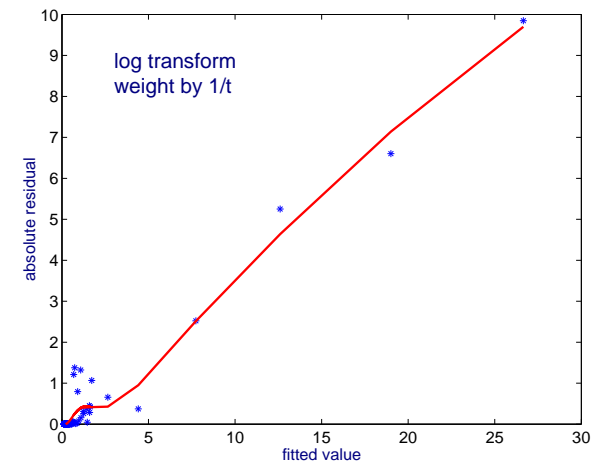
Slide 35



Slide 34



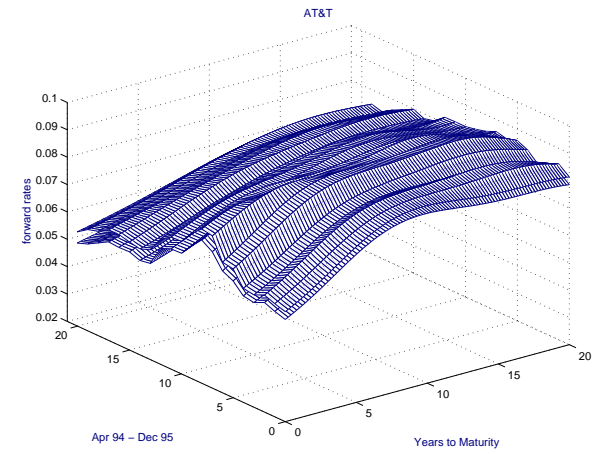
Slide 36



Slide 37

- 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

Slide 39



Slide 38

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

- $\alpha_0 + \alpha_1 t + \alpha_2 t^2$ 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

Slide 40

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

The PLS estimator is the solution to

Slide 41

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

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

Asymptotics: $\lambda \rightarrow 0$

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

Slide 42

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

where

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

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

Slide 43

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

- the large sample variance formula is

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

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

Slide 44

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.

Slide 45

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](#)

Slide 46

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](#)