Time Series and Forecasting
Lecture 4
NonLinear Time Series

Bruce E. Hansen

THE UNIVERSITY of WISCONSIN MADISON

Summer School in Economics and Econometrics
University of Crete
July 23-27, 2012

Today's Schedule

- Density Forecasts
- Threshold Regression Models
- Nonparametric Regression Models

Density Forecasts

- The conditional distribution is

$$F_t(y) = P\left(y_{t+1} \leq y \mid I_t\right)$$

- The conditional density is

$$f_t(y) = \frac{d}{dy} P\left(y_{t+1} \leq y \mid I_t\right)$$

- Density plots are useful summaries of forecast uncertainty
- May also be useful as inputs for other purposes

Density Forecasts

$$\begin{aligned} y_{t+1} &= \mu_t + \sigma_t \varepsilon_{t+1} \\ \mu_t &= E\left(y_{t+1} | I_t\right) \\ \sigma_t^2 &= \operatorname{var}\left(\varepsilon_{t+1} | I_t\right) \end{aligned}$$

- Assume ε_{t+1} is independent of I_t, with density $f^{\varepsilon}(u) = \frac{d}{dy} F^{\varepsilon}(u)$
- Forecast density for y_{n+1}

$$f_n(y) = \frac{1}{\sigma_n} f^{\varepsilon}\left(\frac{y - \mu_n}{\sigma_n}\right)$$

Normal Error Model

- Assume $\varepsilon_{t+1} \sim N(0,1)$, then $f^{\varepsilon}(u) = \phi(u)$

$$\widehat{f}_n(y) = \frac{1}{\widehat{\sigma}_n}\phi\left(\frac{y - \widehat{\beta}'\mathbf{x}_n}{\widehat{\sigma}_n}\right)$$

- Probably should not be used
  - Contains no information beyond and $\widehat{\sigma}_t$

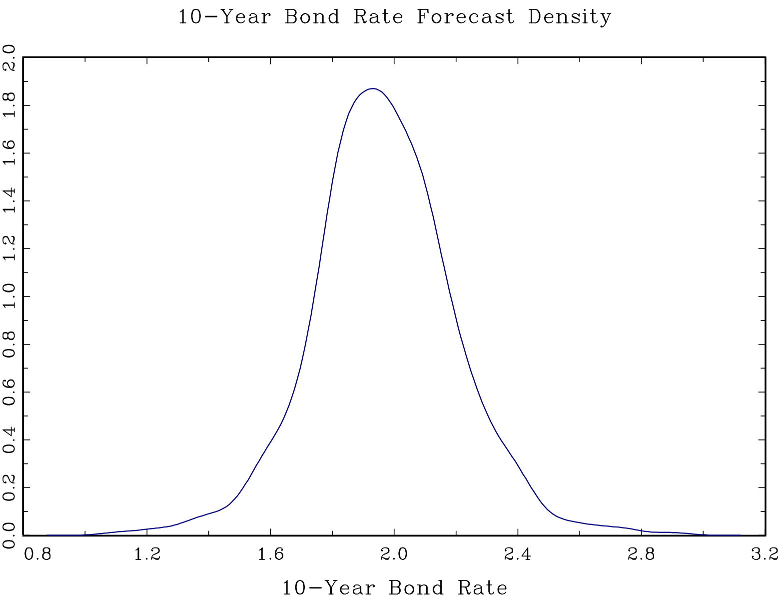
Nonparametric Density Forecast

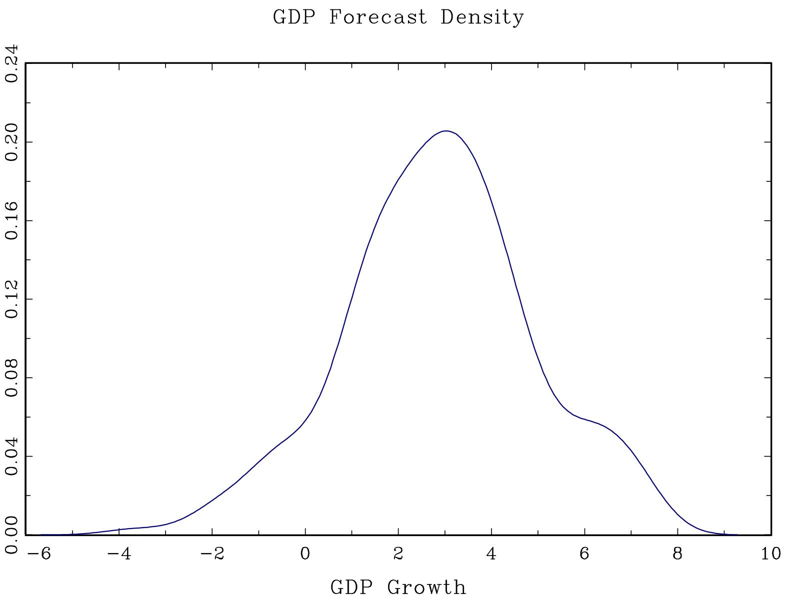
- We can estimate $\widehat{f}^{\varepsilon}(\varepsilon)$ from the normalized residuals $\widehat{\varepsilon}_{t+1}$ using a standard kernel estimator.
  - Discuss this shortly
- Then the forecast density for y_{n+1} is

$$\widehat{f}_n(y) = \frac{1}{\widehat{\sigma}_n}\widehat{f}^{\varepsilon}\left(\frac{y - \widehat{\boldsymbol{\beta}}'\mathbf{x}_n}{\widehat{\sigma}_n}\right)$$

Examples:

- Interest Rate
- GDP Nowcast


10−Year Bond Rate Forecast Density
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
0.8
1.2
1.6
2.0
2.4
2.8
3.2
10−Year Bond Rate


GDP Forecast Density
0.00
0.04
0.08
0.12
0.16
0.20
0.24
−6
−4
−2
0
2
4
6
8
10
GDP Growth

Nonparametric Density Estimation

- Let X_i be a random variable with density $f(x)$
- Observations $i = 1, ..., n$
- [For example, $\hat{\varepsilon}_{t+1}$ for $t = 0, ..., n-1$.]
- The kernel density estimator of $f(x)$ is

$$\hat{f}(x) = \frac{1}{nb} \sum_{i=1}^{n} k\left(\frac{X_i - x}{b}\right)$$

- where $k(u)$ is a kernel function and b is a bandwidth

Kernel Functions

- A **kernel function** $k(u) : \mathbb{R} \to \mathbb{R}$ is any function which satisfies $\int_{-\infty}^{\infty} k(u)du = 1$.
- A **non-negative** kernel satisfies $k(u) \geq 0$ for all u. In this case, $k(u)$ is a probability density function.
- A **symmetric** kernel function satisfies $k(u) = k(-u)$ for all u.
- The **order** of a kernel, ν, is the first non-zero moment.
  - A standard kernel is non-negative, symmetric, and second-order
  - A kernel is **higher-order kernel** if $\nu > 2$. These kernels will have negative parts and are not probability densities. They are also referred to as **bias-reducing kernels**.

Common Second-Order Kernels

Kernel	**Equation**	$R(k)$	κ_2	*eff*
Uniform	$k_0(u) = \frac{1}{2} 1\left(\lvert u\rvert \leq 1\right)$	$1/2$	$1/3$	1.0758
Epanechnikov	$k_1(u) = \frac{3}{4}\left(1 - u^2\right) 1\left(\lvert u\rvert \leq 1\right)$	$3/5$	$1/5$	1.0000
Biweight	$k_2(u) = \frac{15}{16}\left(1 - u^2\right)^2 1\left(\lvert u\rvert \leq 1\right)$	$5/7$	$1/7$	1.0061
Triweight	$k_3(u) = \frac{35}{32}\left(1 - u^2\right)^3 1\left(\lvert u\rvert \leq 1\right)$	$350/429$	$1/9$	1.0135
Gaussian	$k_\phi(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$	$1/2\sqrt{\pi}$	1	1.0513

Choice of Kernel

- Not as important as bandwidth
- Epanechnikov (quadratic) is optimal for minimizing IMSE of $\widehat{f}(x)$
- Gaussian is convenient as it is infinitely smooth and has positive support everywhere
  - I am using Gaussian here

Kernel Density Estimator

- $\widehat{f}(x) = \dfrac{1}{nb} \sum_{i=1}^{n} k\left(\dfrac{X_i - x}{b}\right)$
- $\int_{-\infty}^{\infty} \widehat{f}(x) dx = \int_{-\infty}^{\infty} \frac{1}{n} \sum_{i=1}^{n} \dfrac{1}{b} k\left(\dfrac{X_i - x}{b}\right) dx =$
  $\frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \dfrac{1}{b} k\left(\dfrac{X_i - x}{b}\right) dx = \frac{1}{n} \sum_{i=1}^{n} 1 = 1$
  since by the change of variables $u = (X_i - x)/h$
  $\int_{-\infty}^{\infty} \dfrac{1}{b} k\left(\dfrac{X_i - x}{b}\right) dx = \int_{-\infty}^{\infty} k(u)\, du = 1.$
- Thus $\widehat{f}(x)$ is a density

First Moment

$$\begin{aligned}
\int_{-\infty}^{\infty} x\widehat{f}(x)dx &= \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty} x\frac{1}{b}k\left(\frac{X_i - x}{b}\right)dx \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty}(X_i + uhb)\,k\,(u)\,du \\
&= \frac{1}{n}\sum_{i=1}^{n} X_i \int_{-\infty}^{\infty} k\,(u)\,du + \frac{1}{n}\sum_{i=1}^{n} b\int_{-\infty}^{\infty} uk\,(u)\,du \\
&= \frac{1}{n}\sum_{i=1}^{n} X_i
\end{aligned}$$

the sample mean of the X_i.

Second Moment

$$
\begin{aligned}
\int_{-\infty}^{\infty} x^2 \widehat{f}(x)dx &= \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty} x^2 \frac{1}{b} k\left(\frac{X_i - x}{b}\right) dx \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty} (X_i + ub)^2 k(u)\, du \\
&= \frac{1}{n}\sum_{i=1}^{n} X_i^2 + \frac{2}{n}\sum_{i=1}^{n} X_i b \int_{-\infty}^{\infty} k(u)du + \frac{1}{n}\sum_{i=1}^{n} b^2 \int_{-\infty}^{\infty} u^2 k(u) \\
&= \frac{1}{n}\sum_{i=1}^{n} X_i^2 + b^2\kappa_2
\end{aligned}
$$

where $\kappa_2 = \int_{-\infty}^{\infty} u^2 k(u)\, du$ (1 in the case of Gaussian)

Kernel Density Variance

$$
\begin{aligned}
\int_{-\infty}^{\infty} x^2 \widehat{f}(x)dx - \left(\int_{-\infty}^{\infty} x \widehat{f}(x)dx \right)^2 &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 + b^2\kappa_2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i \right)^2 \\
&= \widehat{\sigma}^2 + b^2\kappa_2
\end{aligned}
$$

where $\widehat{\sigma}^2$ is the sample variance of X_i

In the case of the normalized residuals $\widehat{\varepsilon}_{t+1}$, which have mean zero and sample variance 1, and using a Gaussian kernel:

$\widehat{f}^{\varepsilon}(u)$ has

- A mean of zero
- A variance of $1 + b^2$

Numerical Implementation for Forecast Density

- Pick a set of grid points for ε, e.g. $u_1, ..., u_G$
- For each ε on grid, evaluate

$$\widehat{f}^{\varepsilon}(\varepsilon) = \frac{1}{nb}\sum_{t=0}^{n-1}\phi\left(\frac{\varepsilon - \widehat{\varepsilon}_{t+1}}{b}\right)$$

  or

$$\widehat{f}_j^{\varepsilon} = \frac{1}{nb}\sum_{t=0}^{n-1}\phi\left(\frac{u_j - \widehat{\varepsilon}_{t+1}}{b}\right)$$

- Set the translated gridpoints $y_j = \widehat{\boldsymbol{\beta}}'\mathbf{x}_n + \widehat{\sigma}_n u_j$, for $j = 1, ..., G$, and

$$\widehat{f}_j = \frac{1}{\widehat{\sigma}_n}\widehat{f}_j^{\varepsilon}$$

  The rescaling is the Jacobian of the transformation from u_j to y_j
- Plot $\widehat{f}_j$ on y-axis against y_j on x-axis. This is a plot of

$$\widehat{f}_n(y) = \frac{1}{\widehat{\sigma}_n}\widehat{f}^{\varepsilon}\left(\frac{y - \widehat{\boldsymbol{\beta}}'\mathbf{x}_n}{\widehat{\sigma}_n}\right)$$

Bias of Kernel Estimator

$$\mathrm{E}\widehat{f}(x) = \mathrm{E}\frac{1}{b}k\left(\frac{X_i - x}{b}\right) = \int_{-\infty}^{\infty} \frac{1}{b}k\left(\frac{z-x}{b}\right) f(z)dz$$

Using the change-of variables $u = (z - x)/b$, this equals

$$\int_{-\infty}^{\infty} k(u) f(x + bu)du$$

Now take a Taylor expansion of $f(x + bu)$ about $f(x)$:

$$f(x + bu) \simeq f(x) + f^{(1)}(x)bu + \frac{1}{2}f^{(2)}(x)b^2u^2$$

Integrating term-by term,
$\int_{-\infty}^{\infty} k(u) f(x + bu)du \simeq$
$\int_{-\infty}^{\infty} k(u) f(x) + f^{(1)}(x)b \int_{-\infty}^{\infty} k(u)\, u + \frac{1}{2}f^{(2)}(x)b^2 \int_{-\infty}^{\infty} k(u)\, u^2du$
$= f(x) + \frac{1}{2}f^{(2)}(x)b^2\kappa_2$

Variance of Kernel Estimator

$$
\begin{aligned}
\operatorname{var}\hat{f}(x) &= \frac{1}{n}\operatorname{var}\frac{1}{b}k\left(\frac{X_i - x}{b}\right) \\
&\simeq \frac{1}{nb^2}\int_{-\infty}^{\infty} k\left(\frac{z-x}{b}\right)^2 f(z)dz \\
&= \frac{1}{nb}\int_{-\infty}^{\infty} k(u)^2 f(x+bu)du \\
&\simeq \frac{f(x)}{nb}\int_{-\infty}^{\infty} k(u)^2 du \\
&= \frac{f(x)R(k)}{nb}
\end{aligned}
$$

where $R(k) = \int_{-\infty}^{\infty} k(u)^2 du$ is called the roughness of the kernel.

Asymptotic MSE of Kernel Estimator

$$\begin{aligned} AMSE(\widehat{f}(x)) &= Bias(\widehat{f}(x))^2 + \mathrm{var}\left(\widehat{f}(x)\right) \\ &= \frac{\kappa_2^2}{4}\left(f^{(2)}(x)\right)^2 b^4 + \frac{f(x)\,R(k)}{nb} \end{aligned}$$

Mean Integrated Squared Error (MISE) of Kernel Estimator

$$\begin{aligned} AMISE &= \int_{-\infty}^{\infty} AMSE(\widehat{f}(x))dx \\ &= \int_{-\infty}^{\infty} \frac{\kappa_2^2}{4}\left(f^{(2)}(x)\right)^2 b^4 dx + \int_{-\infty}^{\infty} \frac{f(x)R(k)}{nb}dx \\ &= \frac{\kappa_2^2}{4}R(f)b^4 + \frac{R(k)}{nb} \end{aligned}$$

where $R(f) = \int_{-\infty}^{\infty}\left(f^{(2)}(x)\right)^2 dx$ is the roughness of $f^{(2)}$

Optimal bandwidth

The AMISE takes the form $Ab^4 + B/nb$
The bandwidth b which minimizes the AMISE is

$$b = \left(\frac{4R(k)}{4\kappa_2^2}\right)^{1/5} R(f)^{-1/5} n^{-1/5}$$

The unknown component is $R(f)^{-1/5}$
The "rougher" is $f(x)$, the larger is $R(f)$ so the optimal b is smaller

Rule of Thumb

- Silverman proposed that we take $f = \phi$ as a baseline (or reference)
- Calculate the optimal bandwidth for this case.
  - The "Rule of Thumb"
- $b = \hat{\sigma} C n^{-1/5}$ where

$$C = 2\left(\frac{\pi^{1/2} 2R(k)}{4!\kappa_2^2}\right)^{1/5}$$

Rule of Thumb Constants

Epanechnikov	2.34
Biweight	2.78
Triweight	3.15
Gaussian	1.06

Plug-in bandwidth Methods

Estimate $\widehat{R}(f)$
Use

$$b = \left(\frac{4R(k)}{4\kappa_2^2}\right)^{1/5} \widehat{R}(f)^{-1/5} n^{-1/5}$$

Cross-Validation for Density Bandwidth

- Mean integrated squared error (MISE). Given b

$$
\begin{aligned}
MISE(b) &= \int \left(\widehat{f}(x) - f(x)\right)^2 dx \\
&= \int \widehat{f}(x)^2 dx - 2\int \widehat{f}(x) f(x) dx + \int f(x)^2 dx
\end{aligned}
$$

- We know the first term, not the second, and the third does not depend on b so we ignore it

First Term

- The first term is

$$\int \widehat{f}(x)^2 dx = \int \left(\frac{1}{bh} \sum_{i=1}^{n} k\left(\frac{X_i - x}{b} \right) \right)^2 dx$$
$$= \frac{1}{n^2 b^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \int k\left(\frac{X_i - x}{b} \right) k\left(\frac{X_j - x}{b} \right) dx$$

- Make the change of variables $u = \dfrac{X_i - x}{h}$,

$$\frac{1}{b} \int k\left(\frac{X_i - x}{b} \right) k\left(\frac{X_i - x}{b} \right) dx = \int k(u)\, k\left(u - \frac{X_i - X_j}{b} \right) du$$
$$= k^*\left(\frac{X_i - X_j}{b} \right)$$

where $k^*(x) = \int k(u)\, k(x-u)\, du$ is the convolution of k with itself.

- If $k(x) = \phi(x)$ then $k^*(x) = 2^{-1/2}\phi(x/\sqrt{2}) = \exp(-x^2/4)/\sqrt{4\pi}$.

The first term is thus

$$\int \widehat{f}(x)^2 dx = \frac{1}{n^2 b^2} \sum_{i=1}^{n} \sum_{j=1}^{n} k^* \left(\frac{X_i - X_j}{b} \right)$$

Second Term

- The second term is -2 times
$$\int \widehat{f}(x) f(x) dx$$
an integral with respect to the density of X_i, or an expectation with respect to X_i
- We can estimate expectations using sample averages, e.g. $\frac{1}{n}\sum_{i=1}^{n} \widehat{f}(X_i)$, but $\widehat{f}$ depends on X_i, so this is biased
- The solution is to use a leave-one-out estimator for $\widehat{f}$,
$$\widehat{f}_{-i}(x) = \frac{1}{(n-1)b}\sum_{j \neq i} k\left(\frac{X_j - x}{b}\right)$$
- Then an unbiased estimate of the second term is
$$\frac{1}{n}\sum_{i=1}^{n} \widehat{f}_{-i}(X_i) = \frac{1}{n(n-1)b}\sum_{i=1}^{n}\sum_{j \neq i} k\left(\frac{X_j - X_i}{b}\right)$$

Cross-Validation Criterion

$$MISE(b) = \int \widehat{f}(x)^2 dx - 2\int \widehat{f}(x) f(x) dx + \int f(x)^2 dx$$

$$CV(b) = \frac{1}{n^2 b^2} \sum_{i=1}^{n} \sum_{j=1}^{n} k^* \left(\frac{X_i - X_j}{b} \right) - \frac{2}{n(n-1)b} \sum_{i=1}^{n} \sum_{j \neq i} k \left(\frac{X_j - X_i}{b} \right)$$

In the case of a Gaussian kernel

$$CV(b) = \frac{1}{n^2 b^2 \sqrt{2}} \sum_{i=1}^{n} \sum_{j=1}^{n} \phi \left(\frac{X_i - X_j}{\sqrt{2} b} \right) - \frac{2}{n(n-1)b} \sum_{i=1}^{n} \sum_{j \neq i} \phi \left(\frac{X_j - X_i}{b} \right)$$

- CV selected bandwidth

$$\widehat{b} = \operatorname{argmin} CV(b)$$

Evaluation

- Form a grid for b
- If b_r is the rule-of-thumb bandwidth, search over $[b_r/3, 3b_r]$ or something similar
- Many authors define the CV bandwidth as the largest local minimizer
- In the end, an eyeball reality check of your estimated density is important.

Theory

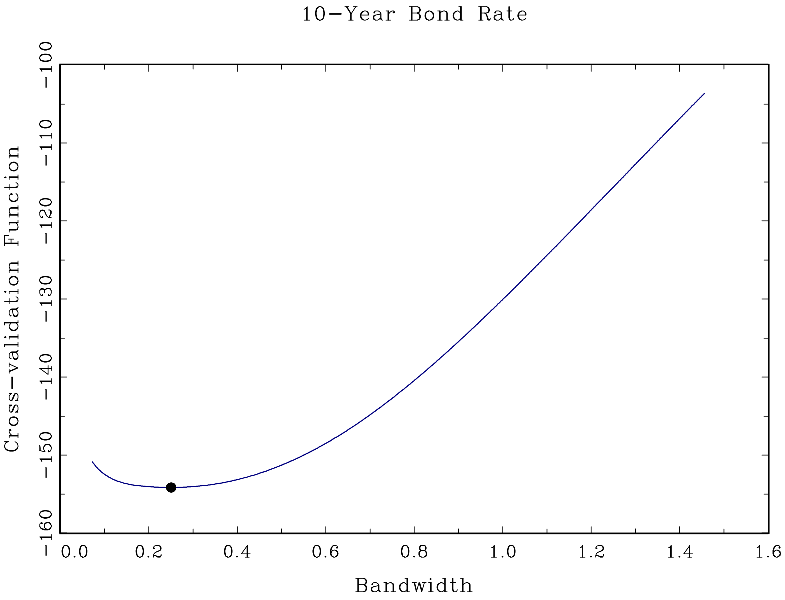
- CV selected bandwidth is consistent
- Let b_0 minimize the AMISE

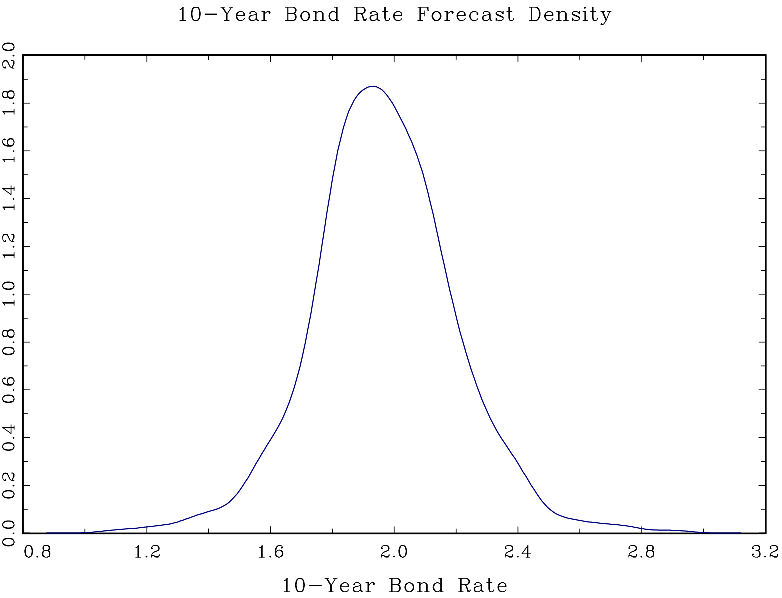
$$\frac{\widehat{b} - b_0}{b_0} \longrightarrow_p 0$$

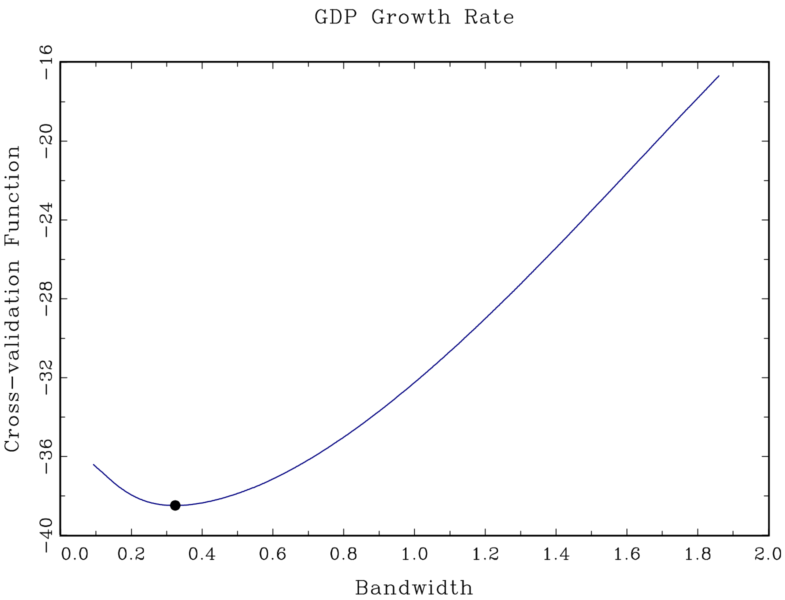
- But the rate of convergence is slow, n^{-10}

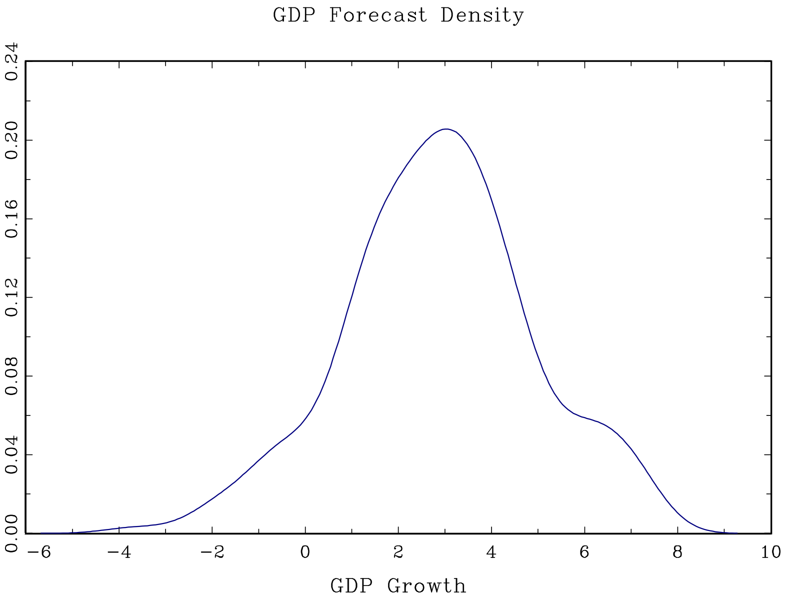
Examples

- 10-Year Bond Rate
- GDP Growth Rate


10−Year Bond Rate
Cross−validation Function
−100
−110
−120
−130
−140
−150
−160
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
Bandwidth


10−Year Bond Rate Forecast Density
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.8
1.2
1.6
2.0
2.4
2.8
3.2
10−Year Bond Rate


GDP Growth Rate
Cross-validation Function
−16
−20
−24
−28
−32
−36
−40
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
Bandwidth


GDP Forecast Density
0.00
0.04
0.08
0.12
0.16
0.20
0.24
−6
−4
−2
0
2
4
6
8
10
GDP Growth

Threshold Models

- A type of nonlinear time series models
- Strong nonlinearity
- Allows for switching effects
- Most typically univariate (for simplicity)

Threshold Models

- Threshold Variable q_t
  - $q_t = 100(\log(GDP_t) - \log(GDP_{t-4})) =$ annual growth
- Threshold γ
- Split regression
  - Coefficients switch if $q_t \leq \gamma$ or $q_t > \gamma$
  - If growth has been above or below the threshold

$$
\begin{aligned}
y_{t+1} &= \boldsymbol{\beta}_1' \mathbf{x}_t 1\left(q_t \leq \gamma\right) + \boldsymbol{\beta}_2' \mathbf{x}_t 1\left(q_t > \gamma\right) + e_{t+1} \\
&= \begin{cases} \boldsymbol{\beta}_1' \mathbf{x}_t + e_t & q_t \leq \gamma \\ \boldsymbol{\beta}_2' \mathbf{x}_t + e_t & q_t > \gamma \end{cases}
\end{aligned}
$$

Partial Threshold Model

$$y_{t+1} = \boldsymbol{\beta}_0' \mathbf{z}_t + \boldsymbol{\beta}_1' \mathbf{x}_t 1 (q_t \leq \gamma) + \boldsymbol{\beta}_2' \mathbf{x}_t 1 (q_t > \gamma) + e_{t+1}$$

- Coefficients on $\mathbf{z}_t$ do not switch
- More parsimonious

Estimation

$$y_{t+1} = \boldsymbol{\beta}_0' \mathbf{z}_t + \boldsymbol{\beta}_1' \mathbf{x}_t 1(q_t \le \gamma) + \boldsymbol{\beta}_2' \mathbf{x}_t 1(q_t > \gamma) + e_{t+1}$$

- Least Squares $(\widehat{\boldsymbol{\beta}}_0, \widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_2, \widehat{\gamma})$ minimize sum-of-squared errors
- Equation is non-linear, so NLLS, not OLS
- Simple to compute by concentration method
  - Given γ, model is linear in $\boldsymbol{\beta}$
  - Regressors are $\mathbf{z}_t$, $\mathbf{x}_t 1(q_t \le \gamma)$ and $\mathbf{x}_t 1(q_t > \gamma)$
  - Estimate by least-squares
  - Save residuals, sum of squared errors
  - Repeat for all thresholds γ. Find value which minimizes SSE

Estimation Details

- For a grid on γ (can use sample values of q_t)
  - Define dummy variables $d_{1t}(\gamma) = 1\,(q_t \leq \gamma)$ and $d_{2t}(\gamma) = 1\,(q_t > \gamma)$
  - Define interaction variables $\mathbf{x}_{1t}(\gamma) = \mathbf{x}_t d_{1t}(\gamma)$ and $\mathbf{x}_{2t}(\gamma) = \mathbf{x}_t d_{2t}(\gamma)$
  - Regress y_{t+1} on $\mathbf{z}_t, \mathbf{x}_{1t}(\gamma), \mathbf{x}_{2t}(\gamma)$

$$y_{t+1} = \widehat{\beta}_0' \mathbf{z}_t + \widehat{\beta}_1' \mathbf{x}_{1t}(\gamma) + \widehat{\beta}_2' \mathbf{x}_{2t}(\gamma) + \hat{e}_{t+1}(\gamma)$$

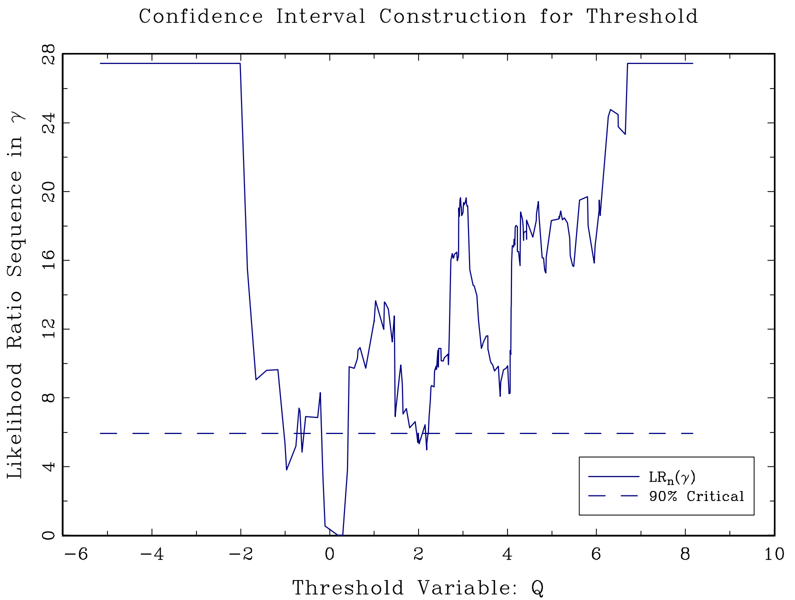
  - Sum of squared errors

$$S(\gamma) = \sum_{t=1}^{n} \hat{e}_{t+1}(\gamma)^2$$

  - Write this explicity as a function of γ as the estimates, residuals and SSE vary with γ
- Find $\hat{\gamma}$ which minimizes $S(\gamma)$
  - Useful to view plot of $S(\gamma)$ against γ
- Given $\hat{\gamma}$, repeat above steps to find estimates $(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2)$
- Forecasts made from fitted model

Example: GDP Forecasting Equation

- $q_t = 100(\log(GDP_t) - \log(GDP_{t-4}))$ = annual growth
- Threshold estimate: $\hat{\gamma} = 0.18$
  - Splits regression depend if past year's growth is above or below 0.18% $\approx 0\%$



Confidence Interval Construction for Threshold
Likelihood Ratio Sequence in γ
0
4
8
12
16
20
24
28
−6
−4
−2
0
2
4
6
8
10
Threshold Variable: Q
LRn(γ)
90% Critical

Multi-Step Forecasts

- Nonlinear models (including threshold models) do not have simple iteration method for multi-step forecasts
- Option 1: Specify direct threshold model
- Option 2: Iterate one-step threshold model by **simulation**:

Multi-Step Simulation Method

- Take fitted model

$$y_{t+1} = \widehat{\beta}_0' \mathbf{z}_t + \widehat{\beta}_1' \mathbf{x}_t 1\left(q_t \le \hat{\gamma}\right) + \widehat{\beta}_2' \mathbf{x}_t 1\left(q_t > \hat{\gamma}\right) + \hat{e}_{t+1}$$

- Draw iid errors $\hat{e}^*_{n+1}, ..., \hat{e}^*_{n+h}$ from the residuals $\{\hat{e}_1, ..., \hat{e}_n\}$
- Create $y^*_{n+1}(b), y^*_{n+2}(b), ..., y^*_{n+h}(b)$ forward by simulation
- b indexes the simulation run
- Repeat B times (a large number)
- $\{y^*_{n+h}(b) : b = 1, ..., B\}$ constitute an iid sample from the forecast distribution for y_{n+h}
  - Point forecast $f_{n+h} = \frac{1}{B} \sum_{b=1}^{B} y^*_{n+h}(b)$
  - Interval forecast: α and $1 - \alpha$ quantiles of $y^*_{n+h}(b)$

Testing for a Threshold

- Null hypothesis: No threshold (linearity)
- Null Model: No threshold

$$
\begin{aligned}
y_{t+1} &= \widehat{\beta}_0' \mathbf{z}_t + \widehat{\beta}_1' \mathbf{x}_t + \hat{e}_{t+1} \\
S_0 &= \sum_{t=1}^{n} \hat{e}_{t+1}^2
\end{aligned}
$$

- Alternative: Single Threshold

$$
\begin{aligned}
y_{t+1} &= \widehat{\beta}_0' \mathbf{z}_t + \widehat{\beta}_1' \mathbf{x}_{1t}(\gamma) + \widehat{\beta}_2' \mathbf{x}_{2t}(\gamma) + \hat{e}_{t+1}(\gamma) \\
S_1(\gamma) &= \sum_{t=1}^{n} \hat{e}_{t+1}(\gamma)^2 \\
S_1 &= S_1(\hat{\gamma}) = \min_{\gamma} S_1(\gamma)
\end{aligned}
$$

Threshold F Test

No Threshold against one threshold

$$F(\gamma) = n\left(\frac{S_0 - S_1(\gamma)}{S_1(\gamma)}\right)$$

$$\begin{aligned} F &= n\left(\frac{S_0 - S_1}{S_1}\right) \\ &= \max_{\gamma} F(\gamma) \end{aligned}$$

NonStandard Testing

- Test is non-standard.
- Critical values obtained by simulation or bootstrap
- Fixed Regressor Bootstrap
  - Similar to a bootstrap, a method to simulate the asymptotic null distribution
  - Fix $(\mathbf{z}_t, \mathbf{x}_t, \hat{e}_t)$, $t = 1, ..., n$
  - Let y_t^* be iid $N(0, \hat{e}_t^2)$, $t = 1, ..., n$
  - Estimate the regressions as before

$$\begin{aligned} y_{t+1}^* &= \widehat{\boldsymbol{\beta}}_0^{*\prime}\mathbf{z}_t + \widehat{\boldsymbol{\beta}}_1^{*}\mathbf{x}_t + \hat{e}_{t+1}^* \\ S_0^* &= \sum_{t=1}^{n} \hat{e}_{t+1}^{*2} \end{aligned}$$

$$\begin{aligned} y_{t+1}^* &= \widehat{\boldsymbol{\beta}}_0^{*\prime}\mathbf{z}_t + \widehat{\boldsymbol{\beta}}_1^{*\prime}\mathbf{x}_{1t}(\gamma) + \widehat{\boldsymbol{\beta}}_2^{*\prime}\mathbf{x}_{2t}(\gamma) + \hat{e}_{t+1}^*(\gamma) \\ S_1^*(\gamma) &= \min_{\gamma} \sum_{t=1}^{n} \hat{e}_{t+1}^*(\gamma)^2 \end{aligned}$$

Bootstrap Test Statistics

$$S_1^* = S_1(\hat{\gamma}) = \min_{\gamma} S_1(\gamma)$$

$$F^*(\gamma) = n\left(\frac{S_0^* - S_1^*(\gamma)}{S_1^*(\gamma)}\right)$$

$$\begin{aligned} F^* &= n\left(\frac{S_0^* - S_1^*}{S_1^*}\right) \\ &= \max_{\gamma} F * (\gamma) \end{aligned}$$

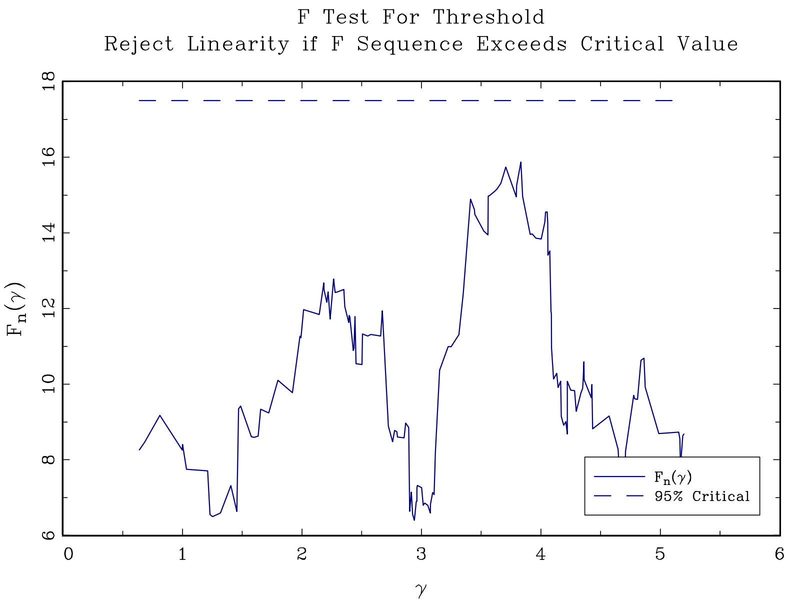
- Repeat this $B \geq 1000$ times.
- Let $F_{01}^*(b)$ denote the b'th value
- Fixed Regressor bootstrap p-value

$$p = \frac{1}{B} \sum_{b=1}^{N} 1 \left(F_{01}^*(b) \geq F_{01} \right)$$

- Fixed Regressor bootstrap critical values are quantiles of empirical distribution of $F_{01}^*(b)$
- Important restriction: Requires serially uncorrelated errors $(h = 1)$

Example: GDP Forecasting Equation

- $q_t = 100(\log(GDP_t) - \log(GDP_{t-4})) =$ annual growth
- Bootstrap p-value for threshold effect: 10.6%


F Test For Threshold
Reject Linearity if F Sequence Exceeds Critical Value
$F_n(\gamma)$
18
16
14
12
10
8
6
0
1
2
3
4
5
6
γ
$F_n(\gamma)$
95% Critical

Inference in Threshold Models

- Threshold Estimate has NonStandard Distribution
- Confidence intervals by inverting F statistic
- F Test: Kknown Threshold against Estimated threshold

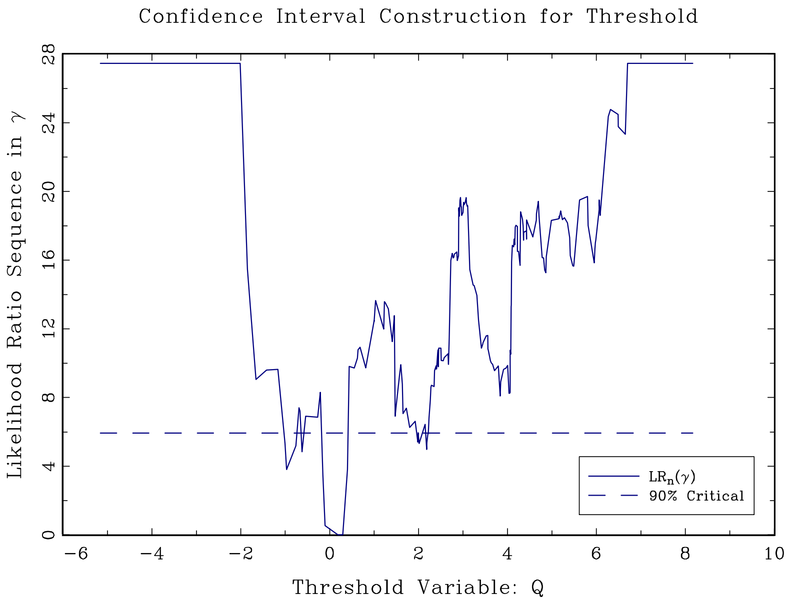
$$LR(\gamma) = n\left(\frac{S_1(\gamma) - S_1}{S_1}\right)$$

  - [Call it $LR(\gamma)$ to distinguish from $F(\gamma)$ from earlier slide.]
- Theory: [Hansen, 2000] $LR(\gamma) \to_d \xi = \max_s [2W(s) - |s|]$
- $P(\xi \leq x) = \left(1 - e^{-x/2}\right)^2$
- Critical values:

$P(\xi \leq c)$	0.80	.90	.95	.99
c	4.50	5.94	7.35	10.59

Confidence Intervals for Threshold

- All γ such that $LR(\gamma) \leq c$ where c is critical value
- Easy to see in graph of $LR(\gamma)$ against γ


Confidence Interval Construction for Threshold
Likelihood Ratio Sequence in γ
0
4
8
12
16
20
24
28
−6
−4
−2
0
2
4
6
8
10
Threshold Variable: Q
$LR_n(\gamma)$
90% Critical

Threshold Estimates

- Estimate: $\hat{\gamma} = 0.18$
- Confidence Interval $= [-1.0, 2.2]$

Inference on Slope Parameters

- Conventional
- As if threshold is known

Threshold Model Estimates

$q_t = 100(\log(GDP_t) - \log(GDP_{t-4}))$

	$q_t \leq 0.18$	$q_t > 0.18$
Intercept	-10.3 (4.6)	-0.23 (1.11)
$\Delta \log(GDP_t)$	0.36 (0.21)	0.16 (0.08)
$\Delta \log(GDP_{t-1})$	-0.22 (0.21)	0.20 (0.09)
Spread_t	1.3 (0.8)	0.71 (0.20)
Default Spread_t	-0.22 (1.26)	-2.3 (0.9)
Housing Starts_t	2.5 (10.6)	4.1 (2.3)
$\text{Building Permits}_t$	7.8 (10.5)	-2.2 (2.0)

NonParametric/NonLinear Time Series Regression

- Optimal point forecast is $g(\mathbf{x}_n)$ where

$$g(\mathbf{x}) = E(y_{t+1} | \mathbf{x}_t = \mathbf{x})$$

  and $\mathbf{x}_t$ are all relevant variables.
- In general, the form of $g(\mathbf{x})$ is unknown and nonlinear
- Linear models used for simplicity, but they are not "true"

NonParametric/NonLinear Time Series Regression

- Model

$$\begin{aligned} y_{t+1} &= g(\mathbf{x}_t) + e_{t+1} \\ E(e_{t+1} | \mathbf{x}_t) &= 0 \end{aligned}$$

- The conditional mean zero restriction holds true by construction
- e_{t+1} not necessarily iid

Additively Separable Model

- $\mathbf{x}_t = (x_{1t}, ..., x_{pt})$

$$g(\mathbf{x}_t) = g_1(x_{1t}) + g_2(x_{2t}) + \cdots + g_p(x_{pt})$$

  Then

$$y_{t+1} = g_1(x_{1t}) + g_2(x_{2t}) + \cdots + g_p(x_{pt}) + e_{t+1}$$

- Greatly reduces degree of nonlinearity
- Useful simplification, but should be viewed as such, not as "true"

Partially Linear Model

- Partition $\mathbf{x}_t = (x_{1t}, \mathbf{x}_{2t})$

$$g(\mathbf{x}_t) = g_1(x_{1t}) + \boldsymbol{\beta}' \mathbf{x}_{2t}$$

- $\mathbf{x}_{2t}$ typically includes dummy variables, controls
- $\mathbf{x}_{1t}$ main variables of importance
- For example, if primary dependence through first lag

$$y_{t+1} = g_1(y_t) + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + e_{t+1}$$

Sieve Models

- For simplicity, suppose x_t is scalar (real-valued)
- WLOG in additively separable and partially linear models
- Approximate $g(x)$ by a sequence $g_m(x)$, $m = 1, 2, ...,$ of increasing complexity
- Linear sieves

$$g_m(x) = Z_m(x)'\boldsymbol{\beta}_m$$

  where $Z_m(x) = (z_{1m}(x), ..., z_{Km}(x))$ are nonlinear functions of x.
- "Series": $Z_m(x) = (z_1(x), ..., z_K(x))$
- "Sieves": $Z_m(x) = (z_{1m}(x), ..., z_{Km}(x))$

Polynomial (power series)

- $z_j(x) = x^j$

$$g_m(x) = \sum_{j=1}^{p} \beta_j x^j$$

- Stone-Weierstrass Theorem: Any continuous function $g(x)$ can be arbitrarily well approximated on a compact set by a polynomial of sufficiently high order
  - For any $\varepsilon > 0$ there exists coefficients p and β_j such that $\mathcal{X}$

$$\sup_{x \in \mathcal{X}} |g_m(x) - g(x)| \leq \varepsilon$$

- Runge's phenomenon:
  - Polynomials can be poor at interpolation (can be erratic)

Splines

- Piecewise smooth polynomials
- Join points are called **knots**
- Linear spline with one knot at τ

$$g_m(x) = \begin{cases} \beta_{00} + \beta_{01}(x - \tau) & x < \tau \\ \beta_{10} + \beta_{11}(x - \tau) & x \geq \tau \end{cases}$$

- To enforce continuity, $\beta_{00} = \beta_{10}$,

$$g_m(x) = \beta_0 + \beta_1 (x - \tau) + \beta_2 (x - \tau) 1 (x \geq \tau)$$

or equivalently

$$g_m(x) = \beta_0 + \beta_1 x + \beta_2 (x - \tau) 1 (x \geq \tau)$$

Quadratic Spline with One Knot

$$g_m(x) = \begin{cases} \beta_{00} + \beta_{01}(x-\tau) + \beta_{02}(x-\tau)^2 & x < \tau \\ \beta_{10} + \beta_{11}(x-\tau) + \beta_{12}(x-\tau)^2 & x \geq \tau \end{cases}$$

- Continuous if $\beta_{00} = \beta_{10}$
- Continuous first derivative if $\beta_{01} = \beta_{11}$
- Imposing these constraints

$$g_m(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 (x-\tau)^2 1(x \geq \tau).$$

Cubic Spline with One Knot

$$g_m(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 (x - \tau)^3 1 (x \geq \tau)$$

General Case

- Knots at $\tau_1 < \tau_2 < \cdots < \tau_N$

$$g_m(x) = \beta_0 + \sum_{j=1}^{p} \beta_j x^j + \sum_{k=1}^{N} \beta_{p+k} (x - \tau_k)^p 1(x \geq \tau_k)$$

Uniform Approximation

- Stone-Weierstrass Theorem: Any continuous function $g(x)$ can be arbitrarily well approximated on a compact set by a polynomial of sufficiently high order
  - For any $\varepsilon > 0$ there exists coefficients p and β_j such that $\mathcal{X}$

$$\sup_{x \in \mathcal{X}} |g_m(x) - g(x)| \leq \varepsilon$$

- Strengthened Form:
  - if the s'th derivative of $g(x)$ is continuous then the uniform approximation error satisfies

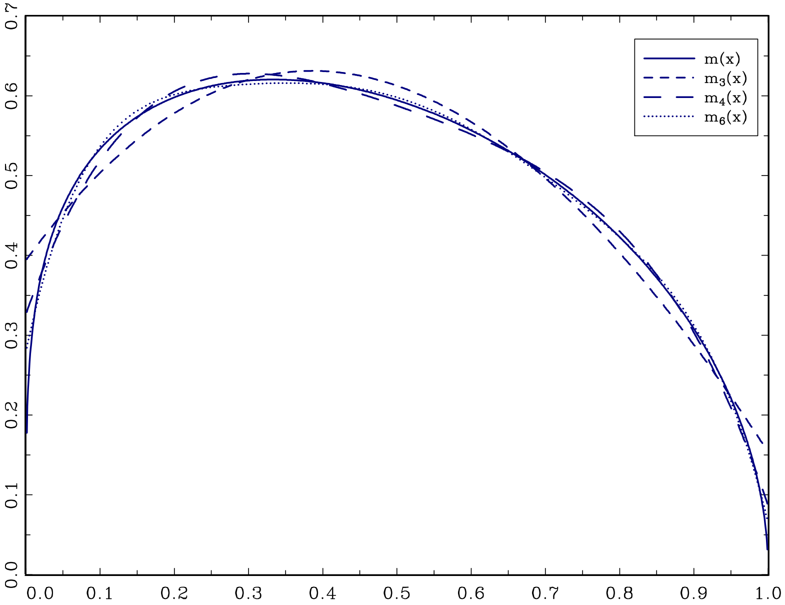
$$\sup_{x \in \mathcal{X}} |g_m(x) - g(x)| = O\left(K_m^{-\alpha}\right)$$

where K_m is the number of terms in $g_m(x)$

- This holds for polynomials and splines
- Runge's phenomenon:
  - Polynomials can be poor at interpolation (can be erratic)

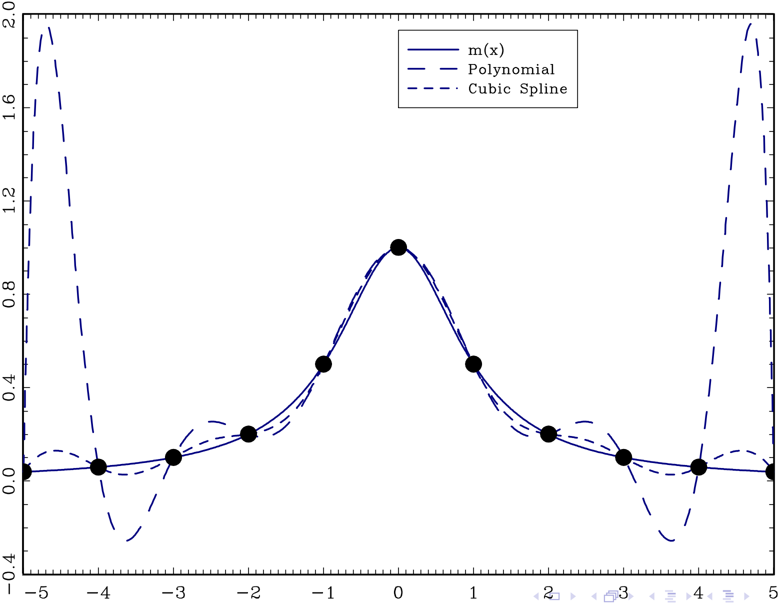
Illustration

- $g(x) = x^{1/4}(1-x)^{1/2}$
- Polynomials of order $K = 3$, $K = 4$, and $K = 6$
- Cubic splines are quite similar



0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
m(x)
$m_3(x)$
$m_4(x)$
$m_6(x)$

Runge's Phenomenon



Placement of Knots

- If support of x is $[0,1]$, typical to set $\tau_j = j/(N+1)$
- If support of x is $[a,b]$, can set $\tau_j = a + (b-a)/(N+1)$
- Alternatively, can set τ_j to equal the $j/(n+1)$ quantile of the distribution of x

Estimation

- Fix number and location of knots
- Estimate coefficients by least-squares
- Quadratic spline

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \sum_{k=1}^{N} \beta_{2+k} (x - \tau_k)^2 1(x \geq \tau_k) + e$$

- Linear model in x, x^2, $(x - \tau_1)^2 1(x \geq \tau_1), ..., (x - \tau_N)^2 1(x \geq \tau_N)$

Selection of Number of Knots

- Model selection
- Pick N to minimize Cross-validation function
- CV is an estimate of
  - MSFE
  - IMSE (integrated mean-squared error)
- CV selection (and combination) is asymptotically optimal for minimization of the MSFE and IMSE

Example: GDP Growth

- y_t =GDP Growth
- x_t =Housing Starts
- Partially Linear Model

$$y_{t+1} = g(x_t) + \beta_1 y_{t-1} + \beta_2 y_{t-2} + e_{t+1}$$

- Polynomial
- Cubic Spline

CV Selection

Polynomial in Housing Starts

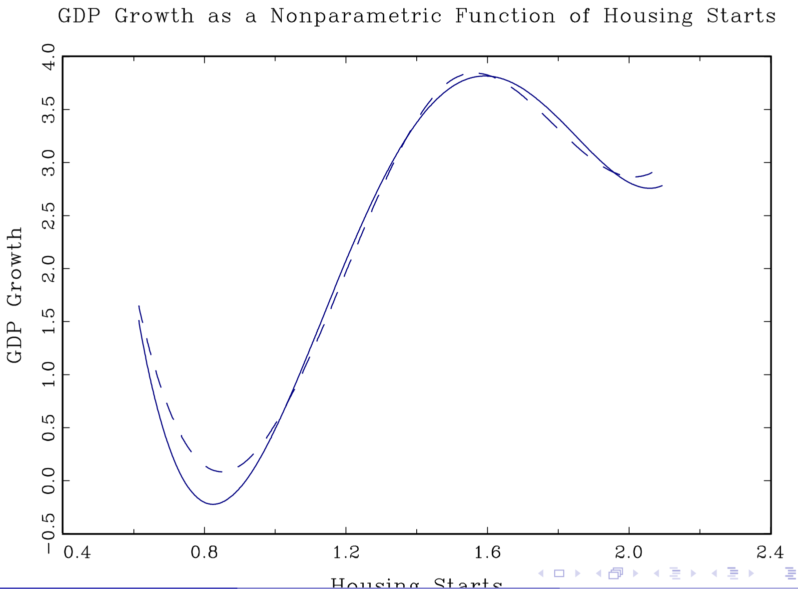
p	1	2	3	4	5	6
CV	10.4	10.5	10.6	9.9	10.0	10.0

Cubic Spline in Housing Starts

N	1	2	3	4	5	6
CV	9.97	10.0	10.0	10.0	10.1	10.2

Best fitting regression is quartic polynomial $(p = 4)$
Cubic spline with 1 knot is close

Polynomial=solid line
Cubic Spline=dashed line



Estimated Cubic Spline

Knot=1.5

	$\widehat{\beta}$	$s(\widehat{\beta})$
Intercept	29	(8)
Δy_t	0.18	(0.08)
Δy_{t-1}	0.10	(0.08)
HS_t	-86	(26)
HS_t^2	79	(23)
HS_t^3	-22	(6)
$(HS_t - 1.5)^2 1(HS_t > 1.5)$	43	(13)

New Example: Long and Short Rates

- Bi-variate model of Long (10-year) and short (3-month) bond rates
- Key variable: Spread: Long-Short
- R_t =Long Rate
- r_t =Short Rate
- $Z_t = R_r - r_t$ =Spread
- Model

$$
\begin{aligned}
\Delta R_{t+1} &= \alpha_0 + \alpha_{p_1}(L)\Delta R_t + \beta_{p_1}(L)\Delta r_t + g_1(Z_t) + e_{1t} \\
\Delta r_{t+1} &= \gamma_0 + \gamma_{p_2}(L)\Delta R_t + \delta_{p_2}(L)\Delta r_t + g_2(Z_t) + e_{2t}
\end{aligned}
$$

CV Selection

- Separately for each equation
  - Long Rate and Short Rate
  - Select over number of lags
  - Number of spline terms for nonlinearity in Spread

CV Selection: Long Rate Equation

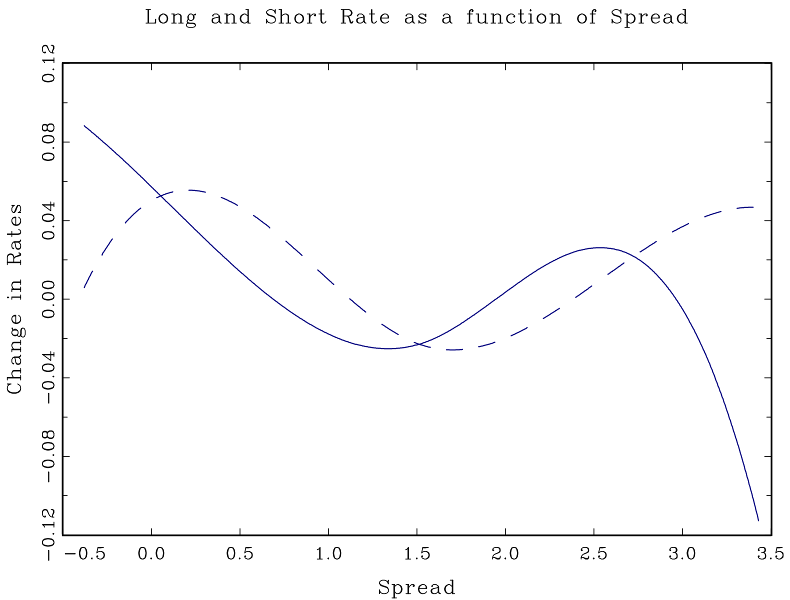
	$p=0$	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=6$
Linear	.0844	.0782	.0760	.0757	.0757	.0766	.0736
Quadratic	.0846	.0781	.0763	.0760	.0760	.0767	.0742
Cubic	.0813	.0794	.0775	.0772	.0771	.0779	.0748
1 Knot	.0821	.0758	.0741	.0739	.0739	.0746	**.0719**
2 Knots	.0820	.0767	.0750	.0747	.0747	.0754	.0724
3 Knots	.0828	.0774	.0758	.0755	.0755	.0762	.0730

Selected Model: $p=6$, Cubic spline with 1 knot at 1.53

CV Selection: Short Rate Equation

	$p = 0$	$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$	$p = 6$
Linear	.206	.183	.181	.186	.189	.193	.186
Quadratic	.203	.178	.177	.181	.185	.187	.183
Cubic	.200	.16979	.172	.176	.179	.181	.179
1 Knot	.198	**.16977**	.172	.176	.179	.180	.179
2 Knots	.200	.172	.174	.178	.182	.183	.181
3 Knots	.201	.171	.174	.179	.182	.183	.181

Selected Model: $p = 1$, Cubic spline with 1 knot at 1.53


Long and Short Rate as a function of Spread
Change in Rates
0.12
0.08
0.04
0.00
−0.04
−0.08
−0.12
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
Spread

Forecasting

- For $h > 1$, need to use forecast simulation
- Simulate R_{n+1}, r_{n+1} forward using iid draws from residuals
- Create time paths
- Take means to estimate point forecasts
- Take quantiles to construct forecats intervals

Assignment

- Construct a nonlinear model to forecast the unemployment rate.
- Use either a threshold or nonparametric model
- Use appropriate methods to select the model and variables
- Make a one-step forecast
- If time, use simulation to create 1 through 12 step forecast distributions. Use this to calculate point forecasts, intervals and a fan chart.