Threshold Autoregressions and NonLinear Autoregressions

Original Presentation: Central Bank of Chile
October 29-31, 2013

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

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

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

Estimation

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

- 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 $\mathbf{z}_t$, $\mathbf{x}_t 1\left(q_t \leq \gamma\right)$ and $\mathbf{x}_t 1\left(q_t > \gamma\right)$
    - 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 $d_{1t}(\gamma) = 1\left(q_t \leq \gamma\right)$ and $d_{2t}(\gamma) = 1\left(q_t > \gamma\right)$
  - 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 $\mathbf{z}_t, \mathbf{x}_{1t}(\gamma), \mathbf{x}_{2t}(\gamma)$

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

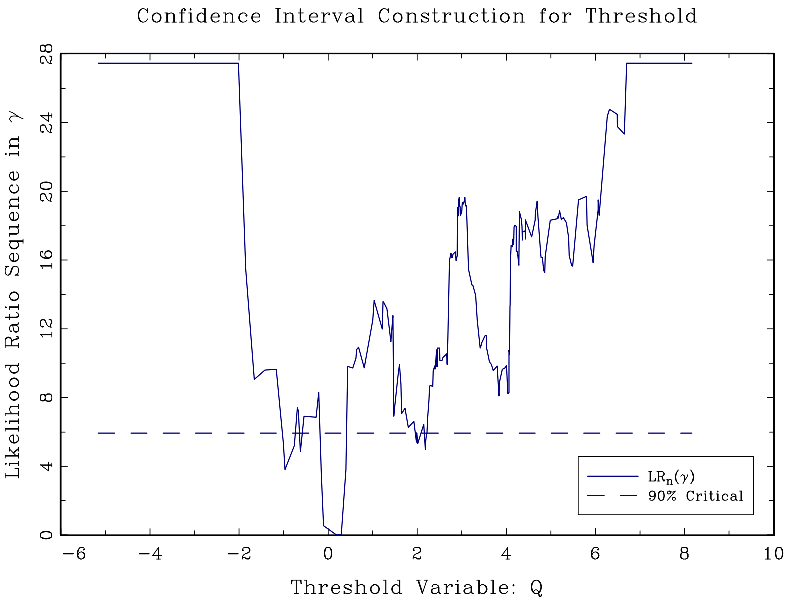
  - Sum of squared errors

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

  - 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

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

- 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

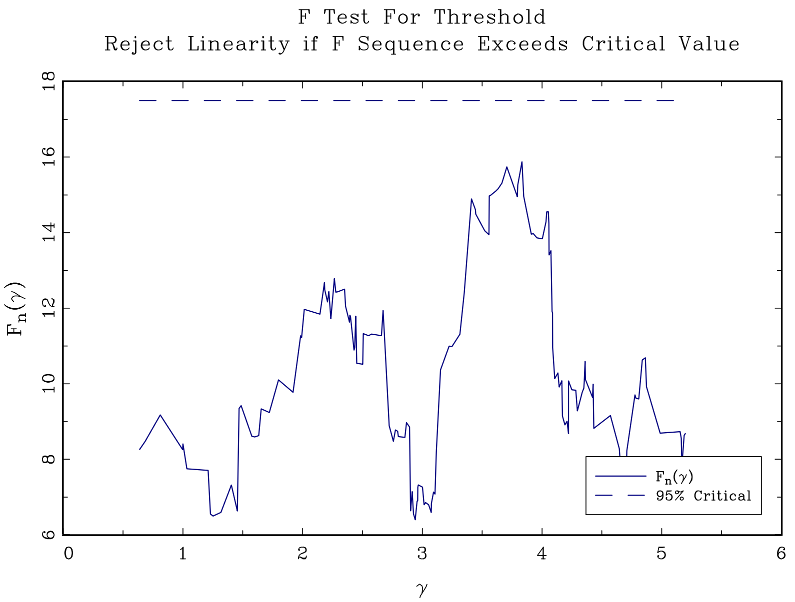
$$\begin{aligned} y_{t+1} &= \widehat{\beta}_0' \mathbf{z}_t + \widehat{\beta}_1' \mathbf{x}_{1t}(\gamma) + \widehat{\beta}_2' \mathbf{x}_{2t}(\gamma) + \hat{e}_{t+1}(\gamma) \\ S_1(\gamma) &= \sum_{t=1}^{n} \hat{e}_{t+1}(\gamma)^2 \\ S_1 &= S_1(\hat{\gamma}) = \min_{\gamma} S_1(\gamma) \end{aligned}$$

NonStandard Testing

- Test is non-standard.
- Similar to Structural Change Tests
- Examine all tests for each fixed threshold
- critical values & p-values obtained by simulation or bootstrap
- Plot sequence of tests; Reject if time plot exceeds critical value

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
$F_n(\gamma)$
γ
18
16
14
12
10
8
6
0
1
2
3
4
5
6
$F_n(\gamma)$
95% Critical

Inference in Threshold Models

- Threshold estimate has nonstandard distribution
- Confidence intervals by inverting statistic constructed from sum-of-squares

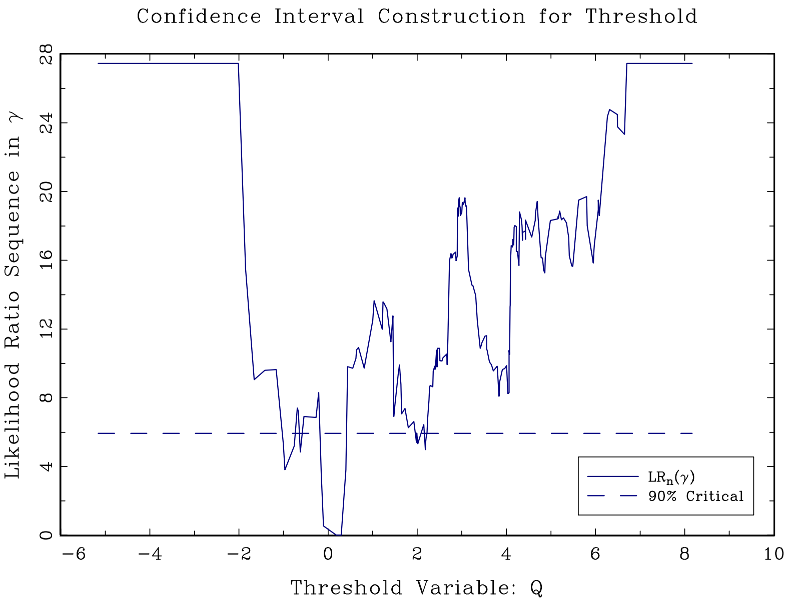
$$LR(\gamma) = n\left(\frac{S_1(\gamma) - S_1}{S_1}\right)$$

- Theory: [Hansen, 2000] $LR(\gamma) \rightarrow_d \xi$
- $P(\xi \leq x) = \left(1 - e^{-x/2}\right)^2$
- 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
Threshold Variable: Q
$LR_n(\gamma)$
90% Critical
0
4
8
12
16
20
24
28
−6
−4
−2
0
2
4
6
8
10

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)
Building Permits$_t$	7.8 (10.5)	-2.2 (2.0)

NonParametric/NonLinear Autoregression

NonParametric/NonLinear Time Series Regression

- Model

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

- $\mathbf{x}_t = (y_{t-1}, y_{t-2}, ..., y_{t-p})$
- or any other variables
- $g(\mathbf{x}_t)$ is arbitrary non-linear function

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

Partially Linear Model

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

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

- $\mathbf{x}_{1t}$ 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

- 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

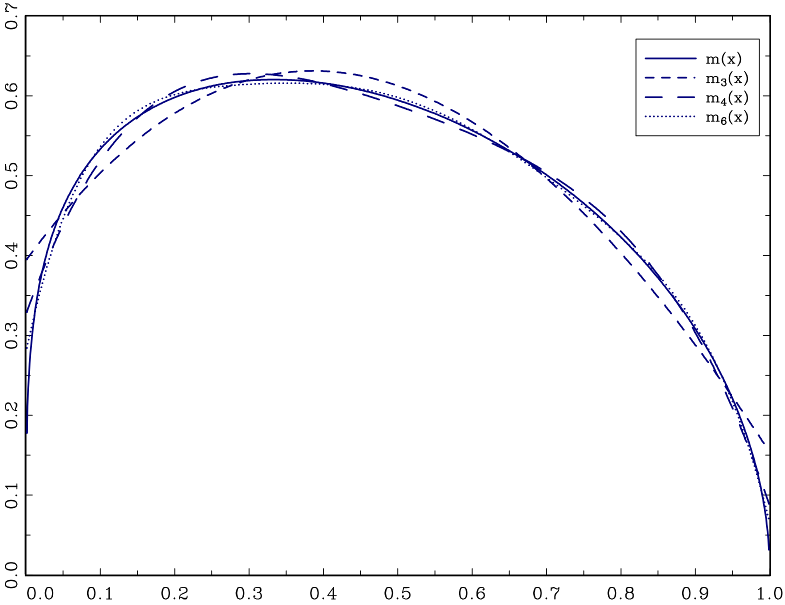
$$\sup_{x \in \mathcal{X}} |g_m(x) - g(x)| = O\left(K_m^{-\alpha}\right)$$

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

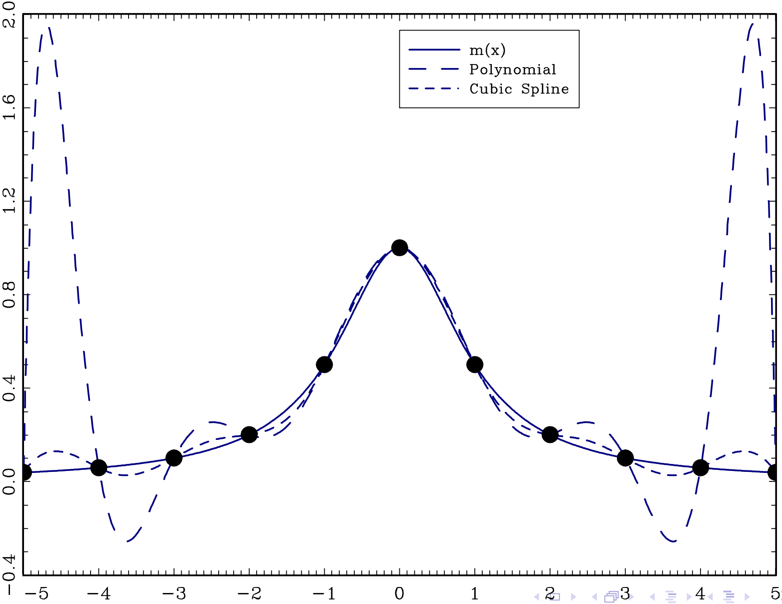
- This holds for polynomials and splines
- Runge's phenomenon:
  - Polynomials can be poor at interpolation (can be erratic)

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


m(x)
m3(x)
m4(x)
m6(x)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

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 AIC
  - Or a similar criterion known as *Cross-Validation* (CV)

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

Model Selection

Polynomial in Housing Starts

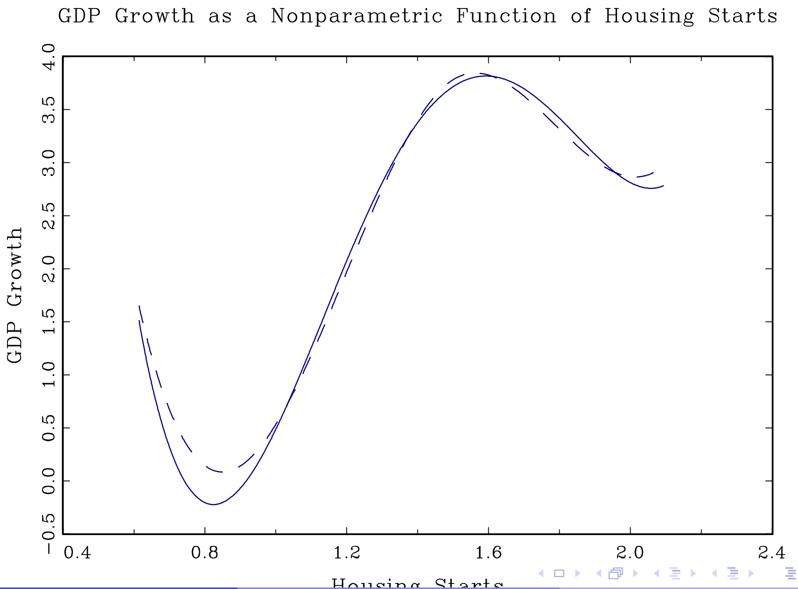
p	1	2	3	4	5	6
CV	10.4	10.5	10.6	9.9	10.0	10.0

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

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

Model 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

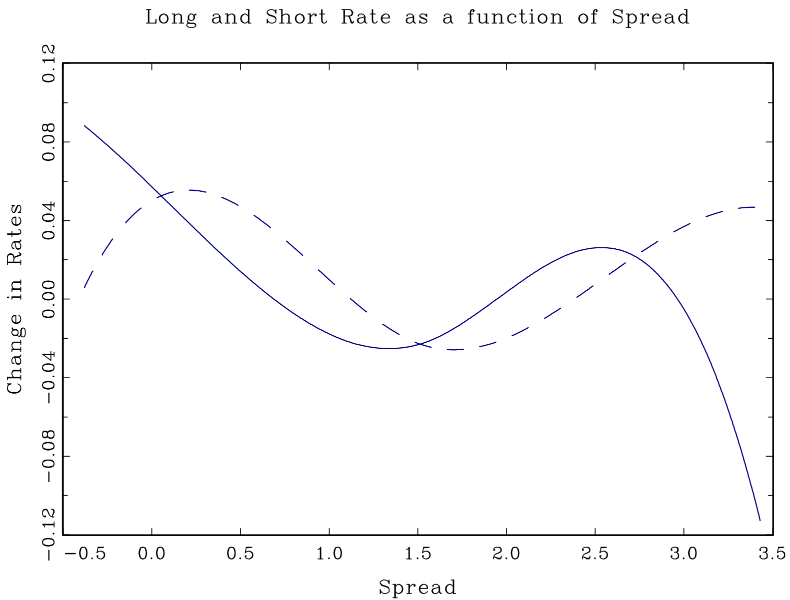
	$p=0$	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=6$
Linear	.0844	.0782	.0760	.0757	.0757	.0766	.0736
Quadratic	.0846	.0781	.0763	.0760	.0760	.0767	.0742
Cubic	.0813	.0794	.0775	.0772	.0771	.0779	.0748
1 Knot	.0821	.0758	.0741	.0739	.0739	.0746	**.0719**
2 Knots	.0820	.0767	.0750	.0747	.0747	.0754	.0724
3 Knots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