Forecast Evaluation

• In our first class, I presented a 12-month extrapolative forecast for the Wisconsin unemployment rate

• How did I do?

• Four realizations:
  – December, January, February, and March
Wisconsin Unemployment Rate - Actual

Unemployment Rate (%)

Historical
90% Forecast Quantile
75% Forecast Quantile
Point Forecast
25% Forecast Quantile
10% Forecast Quantile

Forecast Evaluation

- Actual declined more than point forecast
- 4 realizations within 80% forecast interval
- None within 50% forecast interval
High Dimensional Estimation

• What if you have a situation where the number of regressions $p$ exceeds the number of observations $n$?

• Classic example: gene array data
  – Goal: Determine which gene causes cancer
  – Number of regressors $p = \text{number of genes (5000)}$
  – Number of observations $p = 50$ (or similar)
LASSO

- One solution is LASSO estimation
- Similar idea: LARS, SCAD, Elastic Net
- Idea: Minimize the sum-of-squared errors subject to a penalty based on the sum of the absolute value of the coefficients
LASSO

Model

\[ y_t = \mu + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_p x_{pt} + e_t \]

Minimize sum-of-squared errors plus penalty

\[
\sum_{t=1}^{T} \left( y_t - \mu + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_p x_{pt} \right)^2
\]

\[ + \lambda \sum_{j=1}^{p} |\beta_j| \]

The penalty changes the problem.
Most coefficient estimates are zero.
LASSO and Forecasting

• Lasso very popular in high-dimensional statistics
• I haven’t yet seen Lasso being discussed in economic forecasting
• It is just a matter of time
• Not programmed in Stata
• If interested, I recommend the R package
Software after UW??

- You are unlikely to have access to Stata outside a university environment
  - Some corporations may have a few licenses
  - Non-academic price is expensive
- Excel widely available
  - Often used for regression analysis in corporations
  - Highly limited & clumsy
- R is a viable option
  - Free, open-source
  - Continuously updated
  - Popular among statisticians
  - [http://www.r-project.org/](http://www.r-project.org/)
  - A different style; may need to do more programming
  - Documentation may be limited
NonParametric/NonLinear Time Series Regression

Model

\[ y_{t+1} = g(x_t) + e_{t+1} \]

\[ E(e_{t+1}|x_t) = 0 \]

- The conditional mean zero restriction holds true by construction
- \( e_{t+1} \) not necessarily iid
Additively Separable Model

- \( x_t = (x_{1t}, ..., x_{pt}) \)

\[
g(x_t) = g_1(x_{1t}) + g_2(x_{2t}) + \cdots + g_p(x_{pt})
\]

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

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

- $x_{2t}$ typically includes dummy variables, controls
- $x_{1t}$ 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 $g_m(x)$, $m = 1, 2, \ldots$, of increasing complexity
- Linear sieves
  $$g_m(x) = Z_m(x)' \beta_m$$
  where $Z_m(x) = (z_{1m}(x), \ldots, z_{Km}(x))$ are nonlinear functions of $x$.
- “Series” : $Z_m(x) = (z_1(x), \ldots, z_K(x))$
- “Sieves” : $Z_m(x) = (z_{1m}(x), \ldots, z_{Km}(x))$
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 \( \beta_j \) such that \( \forall x \in \mathcal{X} \)
    \[
    \sup_{x \in \mathcal{X}} |g_m(x) - g(x)| \leq \varepsilon
    \]

- 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 \( \tau \)

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

- 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

\[ g_m(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 (x - \tau)^2 \mathbf{1}(x \geq \tau). \]
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 $\beta_j$ such that
    \[
    \sup_{x \in \mathcal{X}} |g_m(x) - g(x)| \leq \varepsilon
    \]

- Strengthened Form:
  - if the $s'$th derivative of $g(x)$ is continuous then the uniform approximation error satisfies
    \[
    \sup_{x \in \mathcal{X}} |g_m(x) - g(x)| = O(K_m^{-\alpha})
    \]
    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
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 $\tau_j$ 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

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

- Linear model in \( x, x^2, (x - \tau_1)^2 \mathbb{1}(x \geq \tau_1), \ldots, (x - \tau_N)^2 \mathbb{1}(x \geq \tau_N) \)
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

<table>
<thead>
<tr>
<th>$p$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>CV</td>
<td>10.4</td>
<td>10.5</td>
<td>10.6</td>
<td>9.9</td>
<td>10.0</td>
<td>10.0</td>
</tr>
</tbody>
</table>

Cubic Spline in Housing Starts

<table>
<thead>
<tr>
<th>$N$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>CV</td>
<td>9.97</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
<td>10.1</td>
<td>10.2</td>
</tr>
</tbody>
</table>

Best fitting regression is quartic polynomial ($p = 4$)
Cubic spline with 1 knot is close
Polynomial=solid line
Cubic Spline=dashed line

GDP Growth as a Nonparametric Function of Housing Starts
Estimated Cubic Spline

Knot = 1.5

<table>
<thead>
<tr>
<th>Term</th>
<th>$\hat{\beta}$</th>
<th>$s(\hat{\beta})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>29</td>
<td>(8)</td>
</tr>
<tr>
<td>$\Delta y_t$</td>
<td>0.18</td>
<td>(0.08)</td>
</tr>
<tr>
<td>$\Delta y_{t-1}$</td>
<td>0.10</td>
<td>(0.08)</td>
</tr>
<tr>
<td>$HS_t$</td>
<td>-86</td>
<td>(26)</td>
</tr>
<tr>
<td>$HS_t^2$</td>
<td>79</td>
<td>(23)</td>
</tr>
<tr>
<td>$HS_t^3$</td>
<td>-22</td>
<td>(6)</td>
</tr>
<tr>
<td>$(HS_t - 1.5)^2 1 (HS_t &gt; 1.5)$</td>
<td>43</td>
<td>(13)</td>
</tr>
</tbody>
</table>
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
- \( Z_t = R_t - r_t \) = Spread
- Model

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

<table>
<thead>
<tr>
<th></th>
<th>$p = 0$</th>
<th>$p = 1$</th>
<th>$p = 2$</th>
<th>$p = 3$</th>
<th>$p = 4$</th>
<th>$p = 5$</th>
<th>$p = 6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>0.0844</td>
<td>0.0782</td>
<td>0.0760</td>
<td>0.0757</td>
<td>0.0757</td>
<td>0.0766</td>
<td>0.0736</td>
</tr>
<tr>
<td>Quadratic</td>
<td>0.0846</td>
<td>0.0781</td>
<td>0.0763</td>
<td>0.0760</td>
<td>0.0760</td>
<td>0.0767</td>
<td>0.0742</td>
</tr>
<tr>
<td>Cubic</td>
<td>0.0813</td>
<td>0.0794</td>
<td>0.0775</td>
<td>0.0772</td>
<td>0.0771</td>
<td>0.0779</td>
<td>0.0748</td>
</tr>
<tr>
<td>1 Knot</td>
<td>0.0821</td>
<td>0.0758</td>
<td>0.0741</td>
<td>0.0739</td>
<td>0.0739</td>
<td>0.0746</td>
<td><strong>0.0719</strong></td>
</tr>
<tr>
<td>2 Knots</td>
<td>0.0820</td>
<td>0.0767</td>
<td>0.0750</td>
<td>0.0747</td>
<td>0.0747</td>
<td>0.0754</td>
<td>0.0724</td>
</tr>
<tr>
<td>3 Knots</td>
<td>0.0828</td>
<td>0.0774</td>
<td>0.0758</td>
<td>0.0755</td>
<td>0.0755</td>
<td>0.0762</td>
<td>0.0730</td>
</tr>
</tbody>
</table>

Selected Model: $p = 6$, Cubic spline with 1 knot at 1.53
CV Selection: Short Rate Equation

<table>
<thead>
<tr>
<th></th>
<th>$p = 0$</th>
<th>$p = 1$</th>
<th>$p = 2$</th>
<th>$p = 3$</th>
<th>$p = 4$</th>
<th>$p = 5$</th>
<th>$p = 6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>.206</td>
<td>.183</td>
<td>.181</td>
<td>.186</td>
<td>.189</td>
<td>.193</td>
<td>.186</td>
</tr>
<tr>
<td>Quadratic</td>
<td>.203</td>
<td>.178</td>
<td>.177</td>
<td>.181</td>
<td>.185</td>
<td>.187</td>
<td>.183</td>
</tr>
<tr>
<td>Cubic</td>
<td>.200</td>
<td>.16979</td>
<td>.172</td>
<td>.176</td>
<td>.179</td>
<td>.181</td>
<td>.179</td>
</tr>
<tr>
<td>1 Knot</td>
<td>.198</td>
<td><strong>16977</strong></td>
<td>.172</td>
<td>.176</td>
<td>.179</td>
<td>.180</td>
<td>.179</td>
</tr>
<tr>
<td>2 Knots</td>
<td>.200</td>
<td>.172</td>
<td>.174</td>
<td>.178</td>
<td>.182</td>
<td>.183</td>
<td>.181</td>
</tr>
<tr>
<td>3 Knots</td>
<td>.201</td>
<td>.171</td>
<td>.174</td>
<td>.179</td>
<td>.182</td>
<td>.183</td>
<td>.181</td>
</tr>
</tbody>
</table>

Selected Model: $p = 1$, Cubic spline with 1 knot at 1.53
Long and Short Rate as a function of Spread

![Graph showing the relationship between spread and change in rates for long and short rates.](image-url)
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 forecasts intervals