Exponential Smoothing

INSR 260, Spring 2009
Bob Stine
Overview

- Smoothing
- Exponential smoothing
- Model behind exponential smoothing
  - Forecasts and estimates
  - Hidden state model
- Diagnostic: residual plots
- Examples (from Bowerman, Ch 8,9)
  - Cod catch
  - Paper sales
Smoothing

Heuristic

Data = Pattern + Noise

- Pattern is slowly changing, predictable
- Noise may have short-term dependence, but by-and-large is irregular and unpredictable

Idea

Isolate the pattern from the noise by averaging data that are nearby in time.

- Noise should mostly cancel out, revealing the pattern
- Example: moving averages

\[ S_t = \frac{y_{t-w} + \cdots + y_{t-1} + y_t + y_{t+1} + \cdots + y_{t+w}}{2w+1} \]

Example: JMP's spline smoothing uses different weights
Simple Exponential Smooth

- Moving averages have a problem
  - Not useful for prediction:
    Smooth \( s_t \) depends upon observations in the future.
  - Cannot compute near the ends of the data series
- Exponential smoothing is one-sided
  - Average of current and prior values
  - Recent values are more heavily weighted than
  - Tuning parameter \( \alpha = (1-w) \) controls weights \((0 \leq w < 1)\)
- Two expressions for the smoothed value

Weighted average

\[ l_t = \frac{y_t + wy_{t-1} + w^2y_{t-2} + \cdots}{1 + w + w^2 + \cdots} \]

Predictor/Corrector

\[ l_t = \frac{y_t}{1 + w + w^2 + \cdots} + \frac{w(y_{t-1} + wy_{t-2} + \cdots)}{1 + w + w^2 + \cdots} \]

\[ = (1 - w)y_t + wl_{t-1} \]

\[ = \alpha y_t + (1 - \alpha)l_{t-1} \]

\[ = l_{t-1} + \alpha(y_t - l_{t-1}) \]
**Smoothing Constant**

- \( \alpha \) controls the amount of smoothing
  - \( \alpha \approx 0 \) very smooth
  - \( \alpha \approx 1 \) little smoothing

Example (Bowerman): monthly tons, cod

\[
\ell_t = \ell_{t-1} + \alpha (y_t - \ell_{t-1})
\]

Table 6.1
Example: Splines

- Interpolating polynomial
  - always possible to find a polynomial for which \( p(x) = y \) when there is one \( y \) for each \( x \)
  - JMP interactive tool

- Cod catch

---

**Bivariate Fit of Cod (tons) By Time**

- **Smoothing Spline Fit, lambda=0.095749**
  - R-Square: 0.859312
  - Sum of Squares Error: 3702.189

- **Smoothing Spline Fit, lambda=29.62771**
  - R-Square: 0.133897
  - Sum of Squares Error: 22791.47
Example: Exponential Smooth

JMP formula similar to Excel

\[ \ell_t = \ell_{t-1} + \alpha(y_t - \ell_{t-1}) \]

<table>
<thead>
<tr>
<th>Row($) = 1 ⇒ Cod (tons)</th>
</tr>
</thead>
<tbody>
<tr>
<td>else ⇒ Lag[Exp Smth, 1] + alpha * [Cod (tons) - Lag[Exp Smth, 1]]</td>
</tr>
</tbody>
</table>

Which is best?

What value should we use for \( \alpha \)?
Model

Need statistical model to

- Express source of randomness, uncertainty
- Choose an optimal estimate for $\alpha$
- Define predictor and quantify probable accuracy
  - Want to have prediction intervals for exponential smoothing

Latent variable model ("state-space models")

- Assume each observation has mean $L_{t-1}$
  $$y_t = L_{t-1} + \varepsilon_t$$
- Mean values fluctuate over time
  $$L_t = L_{t-1} + \alpha \varepsilon_t$$
- $\varepsilon_t \sim N(0,\sigma^2)$

Discussion

- $L_t$ is the state and is not observed
- If $\alpha = 0$, $L_t$ is constant
- If $\alpha = 1$, $L_t$ is just as variable as the data
Predictions

Model implies a predictor and method for finding prediction intervals

- Observations have mean $L_{t-1}$
  \[ y_t = L_{t-1} + \varepsilon_t \]
- Means fluctuate over time
  \[ L_t = L_{t-1} + \alpha \varepsilon_t \]
- Errors are normally distributed
  \[ \varepsilon_t \sim N(0,\sigma^2) \]

Predictor is constant

- \[ E \ y_{n+1} = L_n \]
- \[ E \ y_{n+2} = L_{n+1} = L_n + \alpha \varepsilon_{n+1} \]
- \[ E \ y_{n+3} = L_{n+2} = L_{n+1} + \alpha \varepsilon_{n+2} = L_n + \alpha (\varepsilon_{n+2} + \varepsilon_{n+2}) = L_n + \alpha (\varepsilon_{n+2} + \varepsilon_{n+2}) \]
- In general, set $\hat{y}_{n+f} = L_n$

Variance of prediction errors grows

- \[ E(y_{n+1}-\hat{y}_{n+1})^2 = E(\varepsilon_{n+1})^2 = \sigma^2 \]
- \[ E(y_{n+f}-\hat{y}_{n+f})^2 = E(\varepsilon_{n+f} + \alpha (\varepsilon_{n+f-1} + \ldots + \varepsilon_{n+1}))^2 = \sigma^2 (1 + (f-1)\alpha^2) \]
Estimating $\alpha$

Model

- Observations have mean $L_{t-1}$
- Mean values fluctuate over time

$y_t = L_{t-1} + \varepsilon_t$
$L_t = L_{t-1} + \alpha \varepsilon_t$

Correspondence

- $l_t$ is our estimate of $L_t$
- $\hat{\alpha}$ is our estimate of $\alpha$ (text uses $\hat{\alpha}$, see page 392)

Estimation

- Like doing least squares but you don’t get to see how well your model captures the underlying state since it is not observed!
- Choose $\hat{\alpha}$ based on forecasting
  - If $L_{t-1}$ were observed, we’d use it to predict $y_t$: it’s the mean of $y_t$
  - Pick $\hat{\alpha}$ to minimize the sum of squared errors, $\Sigma(y_t - l_{t-1})^2$
  - Estimation is not linear in the data

$\varepsilon_t \sim N(0, \sigma^2)$
JMP Results

- Techniques for estimating $\hat{\alpha}$
  - Text illustrates using the Excel solver
  - We’ll use JMP’s time series methods
    - Analyze > Modeling > Time Series

- Simple exponential smoothing
  - Lots of output
  - Results confirm little smooth pattern; near constant

Book “cheats” a little by setting $y_0$ to mean of first 12 rather than smoothing after $y_1$. Finds $\hat{\alpha} \approx 0.03$. 

<table>
<thead>
<tr>
<th>Model Summary</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>DF</td>
<td>22</td>
<td>Stable</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum of Squared Errors</td>
<td>26315.4779</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance Estimate</td>
<td>1196.15809</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>34.5855184</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Akaike’s ‘A’ Information Criterion</td>
<td>232.424394</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schwarz’s Bayesian Criterion</td>
<td>233.559888</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RSquare</td>
<td>-0.174024</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RSquare Adj</td>
<td>-0.174024</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAPE</td>
<td>31.0025784</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAE</td>
<td>230.424394</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hessian is not positive definite.

Parameter Estimates

| Term                                      | Estimate  | Std Error | t Ratio | Prob>| |t| |
|-------------------------------------------|-----------|-----------|---------|-----|-----|-----|
| Level Smoothing Weight                    | 0.00001974| 0         | 0.0000* | 0.0000*|

Forecast

260
280
300
320
340
360
380
400
420
440
460
480
500
0
10
20
30
40
50
Row
Predicted Value

$\hat{\alpha} \approx 0$
Diagnostics

- Sequence plot of residuals
  - One-step ahead prediction errors, $y_t - \hat{y}_t$
  - Normal quantile plot

- No visual pattern remains
  - But we will in a week or so more elaborate diagnostic routines associated with ARIMA models
  - Text discusses tracking methods
Example

- Data are weekly sales of paper towels
  - Goal is to forecast future sales
  - Units of data are 10,000 rolls

- Level appears to change over time, trending down then up.
Exponential Smooth

- Apply simple exponential smoothing
- Model results not very satisfying
  - Value for smoothing parameter, $\hat{\alpha} = 1$ (max allowed)
  - Forecasts are constant
- Motivates alternative smoothing methods

---

Parameter Estimates

| Term                  | Estimate  | Std Error | t Ratio | Prob>|t| |
|-----------------------|-----------|-----------|---------|--------|
| Level Smoothing Weight| 1.0000000 | 0.1110102 | 9.01    | <.0001*|

---

Model Summary

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>DF</td>
<td>118</td>
<td>Stable</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Sum of Squared Errors</td>
<td>143.840613</td>
<td>Invertible</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Variance Estimate</td>
<td>1.21898825</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>1.10407801</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Akaike's 'A' Information Criterion</td>
<td>362.267669</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schwarz's Bayesian Criterion</td>
<td>365.046793</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RSquare</td>
<td>0.93712456</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RSquare Adj</td>
<td>0.93712456</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAPE</td>
<td>18.7016851</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAE</td>
<td>0.86251008</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2LogLikelihood</td>
<td>360.267669</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Summary

- Smoothing
  - locate patterns
- Exponential smoothing
  - uses past
- Model for exponential smoothing
  - latent state
- Diagnostic: residual plots
  - patternless

Discussion
- Desire predictions that are more dynamic
- Extrapolate trends
  - Linear patterns
  - Seasonal patterns