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

$$\boxed{s_t = \frac{y_{t-w} + \cdots + y_{t-1} + y_t + y_{t+1} + \cdots + y_{t+w}}{2w+1}}$$

  - 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 α = (1-w) controls weights (0≤w<1)
- Two expressions for the smoothed value

Weighted average

$$\ell_t = \frac{y_t + w y_{t-1} + w^2 y_{t-2} + \cdots}{1 + w + w^2 + \cdots}$$

Predictor/Corrector

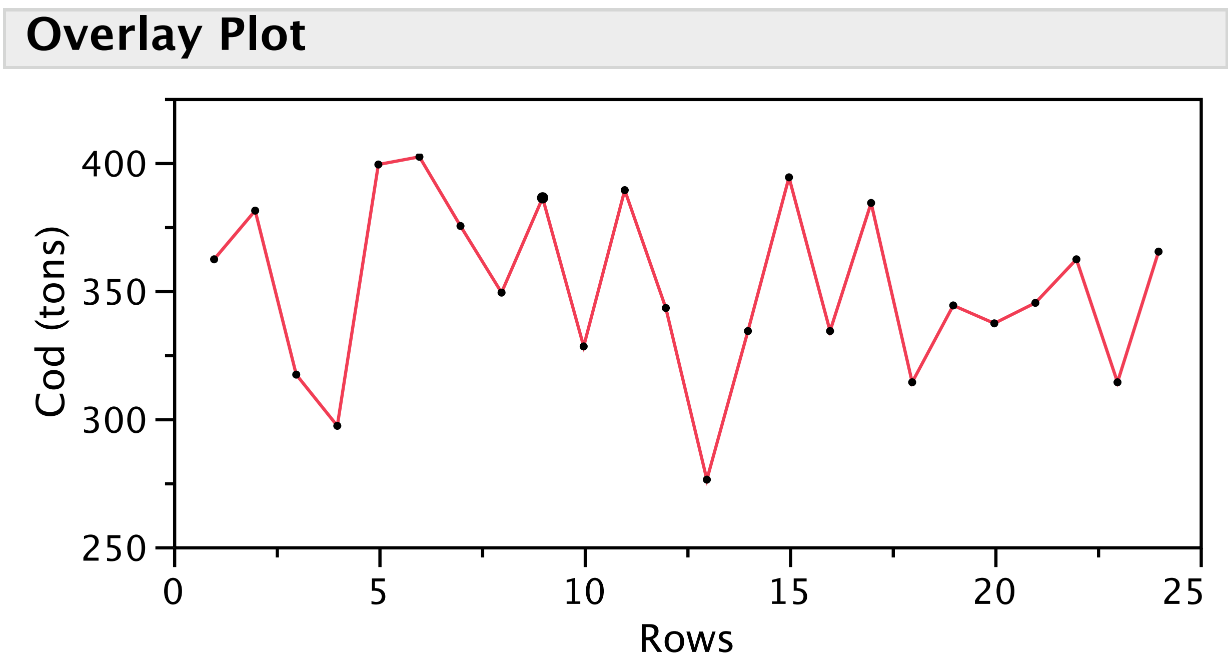
$$\begin{aligned}
\ell_t &= \frac{y_t}{1 + w + w^2 + \cdots} + \frac{w(y_{t-1} + w y_{t-2} + \cdots}{1 + w + w^2 + \cdots} \\
&= (1-w) y_t + w \ell_{t-1} \\
&= \alpha y_t + (1-\alpha) \ell_{t-1} \\
&= \ell_{t-1} + \alpha(y_t - \ell_{t-1})
\end{aligned}$$

Smoothing Constant

- α controls the amount of smoothing
  - $\alpha \approx 0$ very smooth
  - $\alpha \approx 1$ little smoothing

$$\boxed{\ell_t = \ell_{t-1} + \alpha(y_t - \ell_{t-1})}$$

- Example (Bowerman): monthly tons, cod



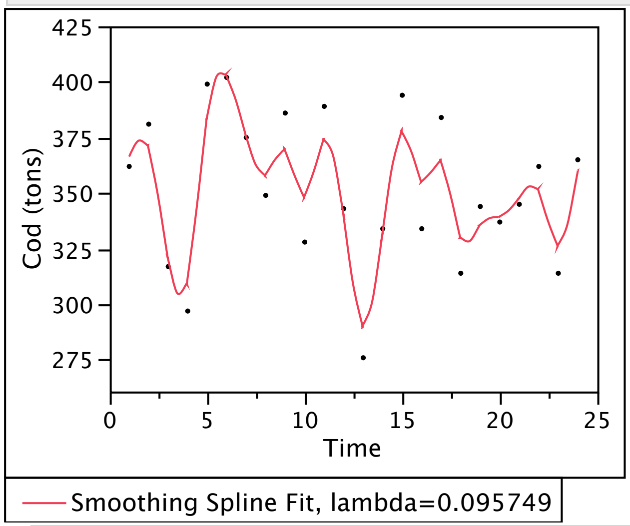
Your impression of the smooth?

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

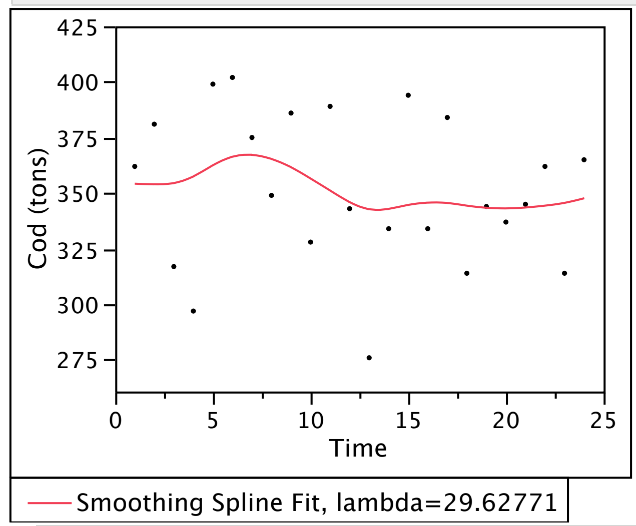
Bivariate Fit of Cod (tons) By Time



Smoothing Spline Fit, lambda=0.095749

R-Square	0.859312
Sum of Squares Error	3702.189

Bivariate Fit of Cod (tons) By Time



Smoothing Spline Fit, lambda=29.62771

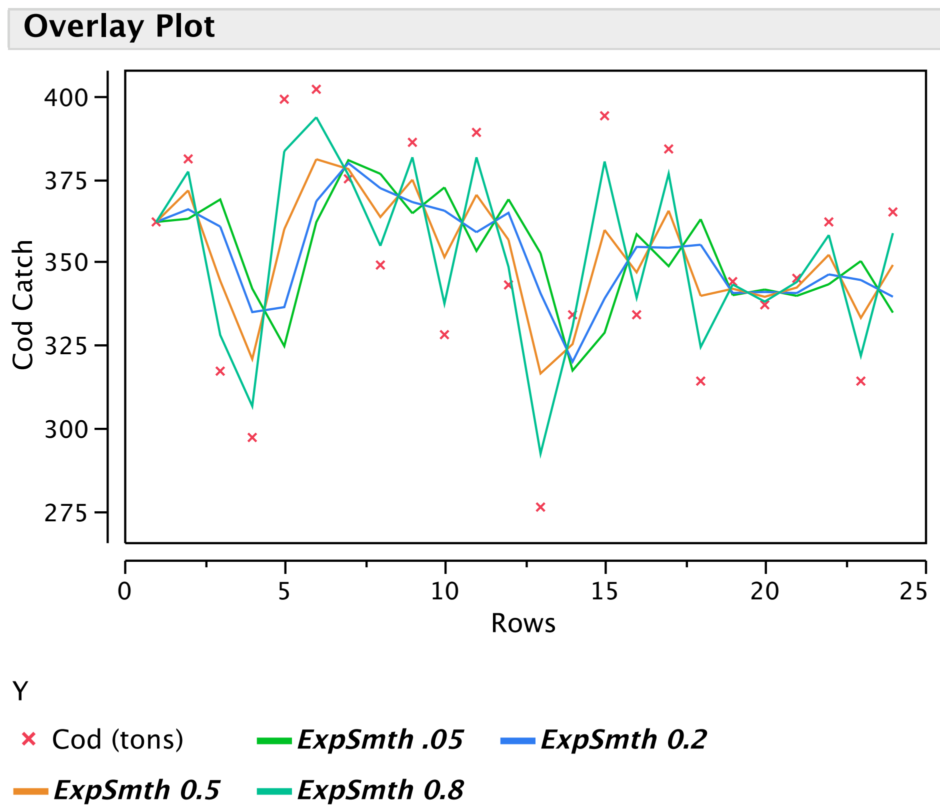
R-Square	0.133897
Sum of Squares Error	22791.47

Example: Exponential Smooth

- JMP formula similar to Excel

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

```
If  Row() == 1  ⇒ Cod (tons)
    else        ⇒ Lag( Exp Smth , 1 ) + alpha * ( Cod (tons) - Lag( Exp Smth , 1 ) )
```



Which is best?

What value should we use for α?

Model

- Need statistical model to
  - Express source of randomness, uncertainty
  - Choose an optimal estimate for α
  - 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 $\qquad \varepsilon_t \sim N(0,\sigma^2)$

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

  - 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
  - $E\ y_{n+1} = L_n$ — $\hat{y}_{n+1} = L_n$
  - $E\ y_{n+2} = L_{n+1} = L_n + \alpha\varepsilon_{n+1}$ — $\hat{y}_{n+2} = L_n$
  - $E\ y_{n+3} = L_{n+2} = L_{n+1}+\alpha\varepsilon_{n+2} = L_n+\alpha(\varepsilon_{n+2}+\varepsilon_{n+2})$ — $\hat{y}_{n+2} = L_n$
  - 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 α

- Model
  - Observations have mean L_{t-1} $\qquad y_t = L_{t-1} + \varepsilon_t \quad \varepsilon_t \sim N(0,\sigma^2)$
  - Mean values fluctuate over time $\qquad L_t = L_{t-1} + \alpha\, \varepsilon_t$
- Correspondence
  - l_t is our estimate of L_t
  - $\hat{a}$ is our estimate of α (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 $\hat{a}$ 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 $\hat{a}$ to minimize the sum of squared errors, $\Sigma(y_t - l_{t-1})^2$
    - Estimation is <u>not</u> linear in the data

JMP Results

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

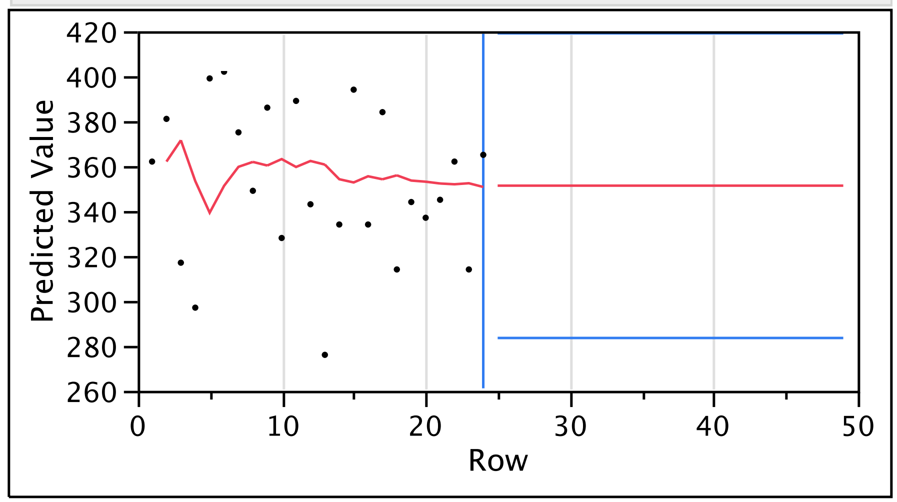
- Simple exponential smoothing
  - Lots of output
  - Results confirm little smooth pattern; near constant

$\hat{a} \approx 0$

Parameter Estimates

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

Forecast



Model Summary

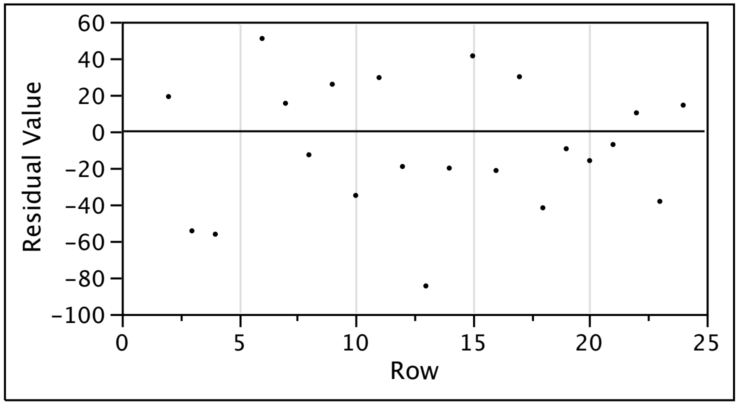
DF	22	Stable	Yes
Sum of Squared Errors	26315.4779	Invertible	Yes
Variance Estimate	1196.15809		
Standard Deviation	34.5855184		
Akaike's 'A' Information Criterion	232.424394		
Schwarz's Bayesian Criterion	233.559888		
RSquare	-0.174024		
RSquare Adj	-0.174024		
MAPE	9.14722694		
MAE	31.0025784		
-2LogLikelihood	230.424394		

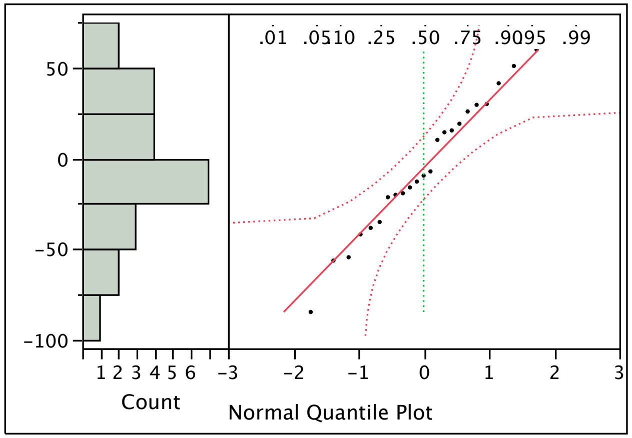
Hessian is not positive definite.

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

Diagnostics

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



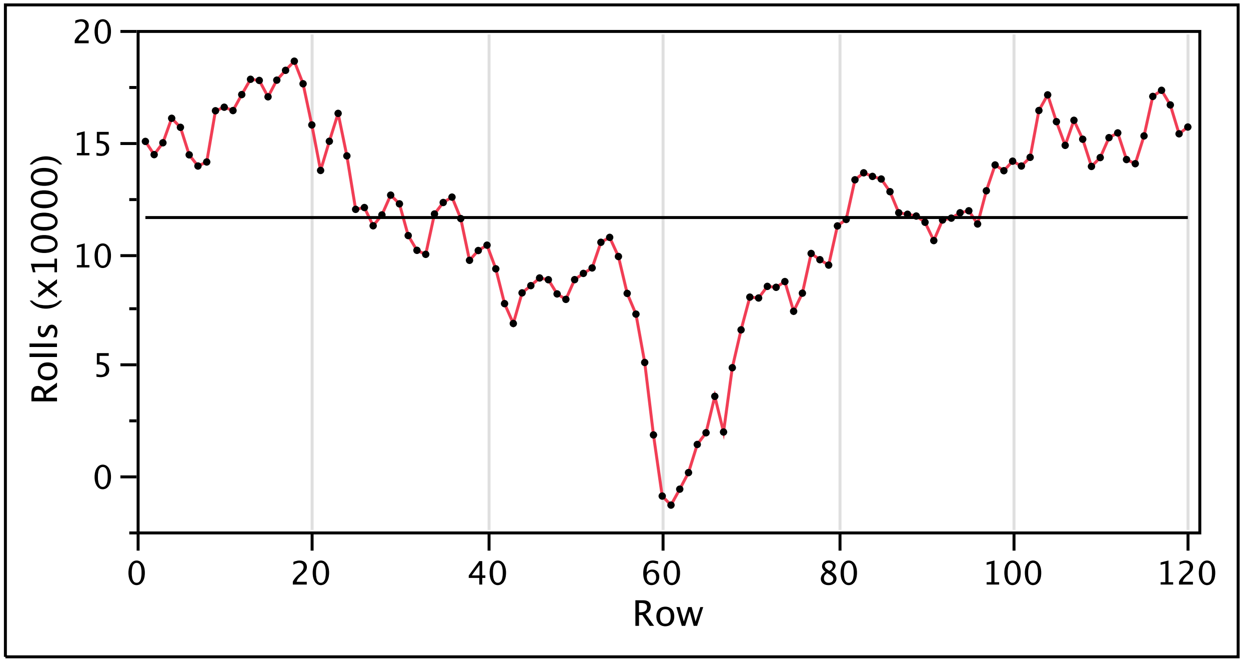


- No visual pattern remains
  - But we will in a week or so more elaborate diagnostic routines associated with ARIMA models
  - Text discusses tracking methods

Example

Table 9.1

- 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

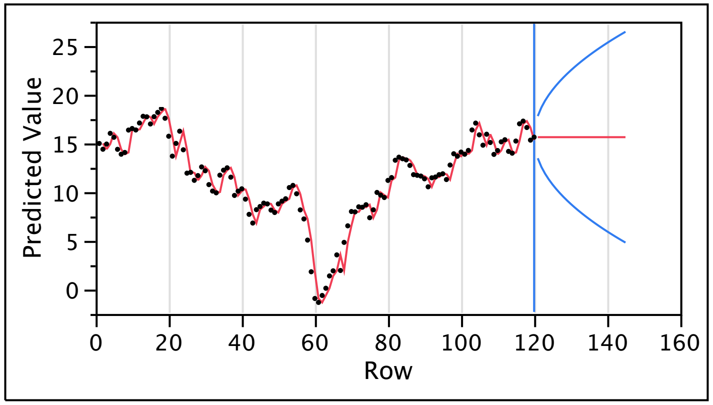
- Apply simple exponential smoothing
- Model results not very satisfying
  - Value for smoothing parameter, â = 1 (max allowed)
  - Forecasts are constant
- Motivates alternative smoothing methods

Parameter Estimates

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

Model Summary

DF	118	Stable	Yes
Sum of Squared Errors	143.840613	Invertible	Yes
Variance Estimate	1.21898825		
Standard Deviation	1.10407801		
Akaike's 'A' Information Criterion	362.267669		
Schwarz's Bayesian Criterion	365.046793		
RSquare	0.93712456		
RSquare Adj	0.93712456		
MAPE	18.7016851		
MAE	0.86251008		
-2LogLikelihood	360.267669		



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