Data Mining and Business Intelligence

Lecture 8: Exponential Smoothing Models

Jing Peng
University of Connecticut

3/12/20

Agenda

- Midterm Survey
- Online class due to coronavirus
- Review of Lecture 7
- Forecasting
- Exponential Smoothing Models
- Applications

Midterm Feedback

- Relate math to real-life cases and explain business insights
- Outline some key points that we should master
- Slow down a bit and spend more time on practices
- Explain concepts more thoroughly and slowly
- Less dig into math
- More concepts than tools. Dive deeper into specific topics
- Provide non-SAS material (e.g., Python)
- More assignments

Stationarity

- What are the three properties of stationary time series?

- Can a stationary time series exhibit trend or seasonality?

Stationarity of AR(1) Model

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \varepsilon_t$$

- Is the model stationary if $|\phi_1| > 1$?

- Is the model stationary if $\phi_1 = 1$?

Multiple AR.R

- Is the model stationary if $\phi_1 = -1$?

- What if $|\phi_1| < 1$ but $\phi_0 > 1$?

Stationarity of MA(1) Model

$$Y_t = \theta_0 + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

- Is the model stationary if $|\theta_1| > 1$?

Multiple MA.R

Which of the following statements are true?

- A random walk model has constant mean
- A random walk model has constant variance
- A random walk with drift model has constant mean
- A random walk with drift model has constant variance

multiple AR.R

Forecasting

Statistical Forecasting — The Math

- Time Series = Signal + Noise

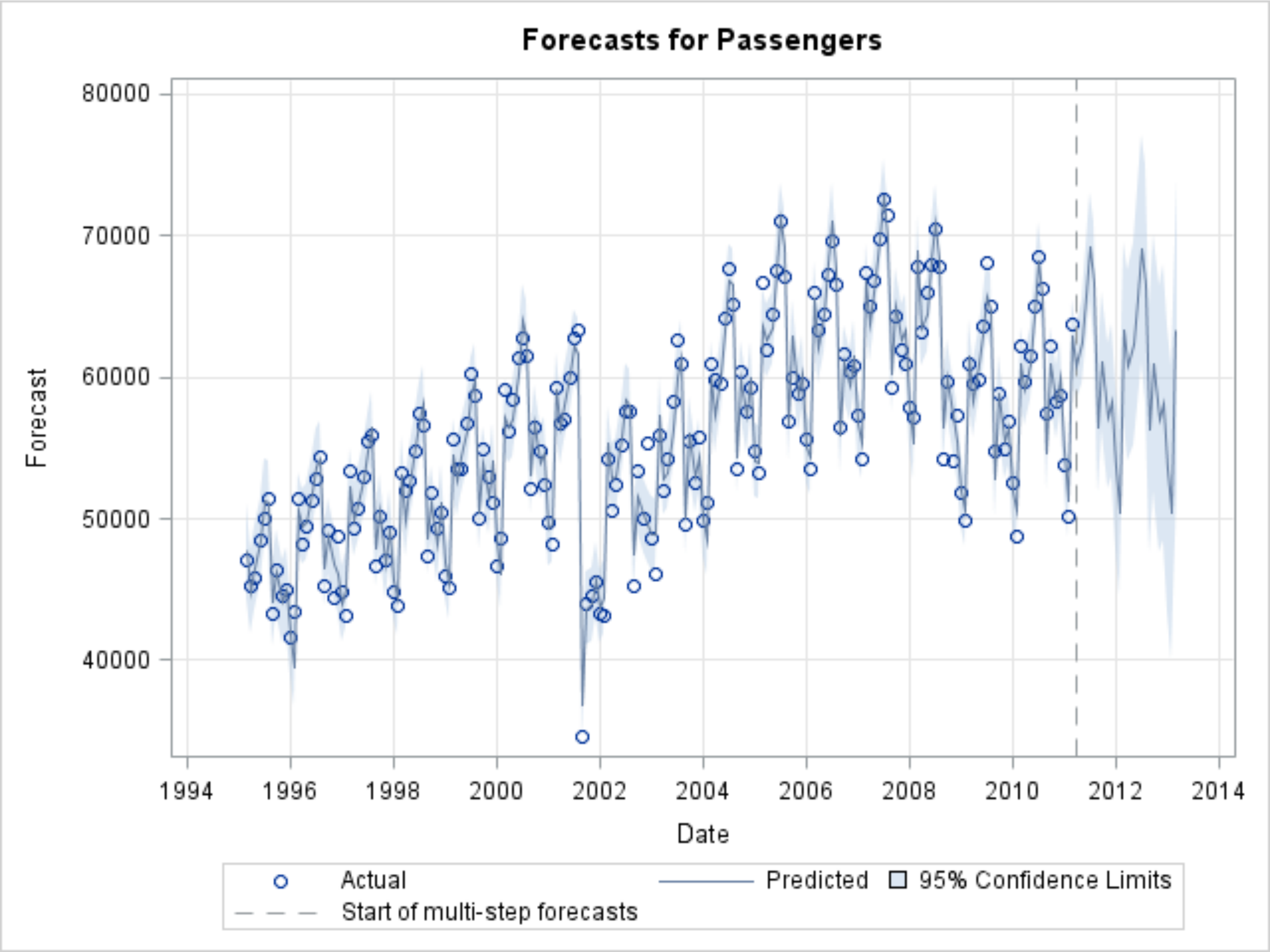
  Systematic Variation   Uncertainty

- Forecast = Extrapolated Signal

- Confidence Interval = Extrapolated Signal +/- Uncertainty

- Example: $Y_t = \phi_0 + \phi_1 Y_{t-1} + \varepsilon_t$

Forecast = Extrapolated Signal



Model Evaluation Metrics

- Root Mean Squared Error

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \widehat{Y_t}\right)^2}$$

- Mean Absolute Percent Error:

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n} | Y_t - \hat{Y}_t | / Y_t$$

- Mean Absolute Error:

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n} | Y_t - \hat{Y}_t |$$

Evaluation Metrics

- Akaike Information Criteria (AIC):

$$\mathrm{AIC} = -2\log(L) + 2k$$

- Schwarz's Bayesian Information Criteria (SBC or BIC):

$$\mathrm{SBC} = -2\log(L) + k\log(n)$$

- L represents the likelihood of the model
- k represents the number of parameters in the model
- n represents number of observations in the data

Evaluation Process

Fit Sample (Fit)

- **Used to estimate model parameters for accuracy evaluation**

Holdout Sample (Accuracy)

- **Used to evaluate model accuracy**

⚠ **Full = Fit + Holdout** data is used to fit a **final** deployment model

Choosing the Holdout Sample

- Rules of thumb
  - Cover at least a complete seasonal period in holdout sample
  - Length of fit sample is 4 times more than number of parameters
  - Holdout sample often contains no more than 25% of the series
  - The holdout sample is always at the end of the series

- No holdout sample and base accuracy on the entire series
  - If unique behavior occurs within the holdout sample
  - If there is insufficient data to fit a model without the holdout sample

Exponential Smoothing Models

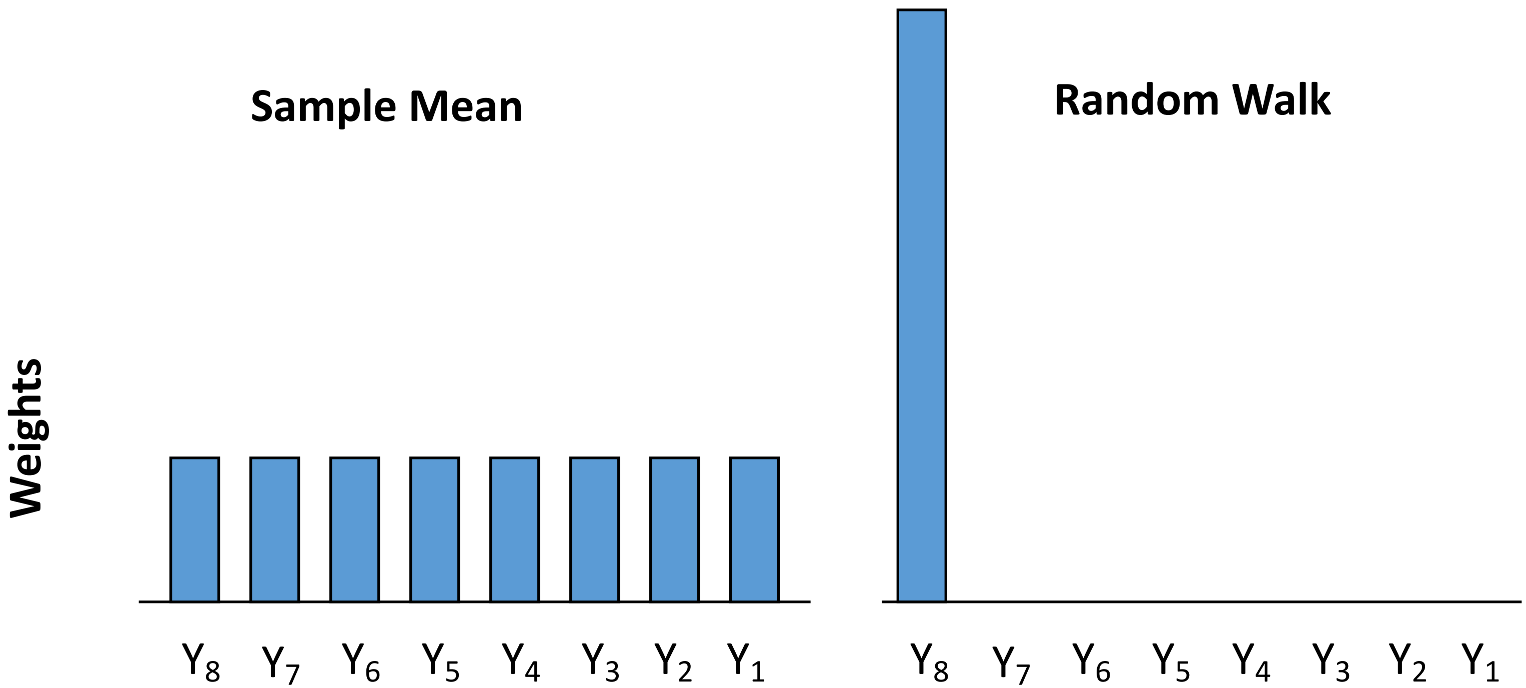
Exponential Smoothing

- Proposed in the late 1950s (Brown, 1959; Holt, 1957; Winters, 1960)

- Produce forecasts using weighted averages of past observations, with the weights decaying exponentially as the observations get older.

- Support data with trend and/or seasonality

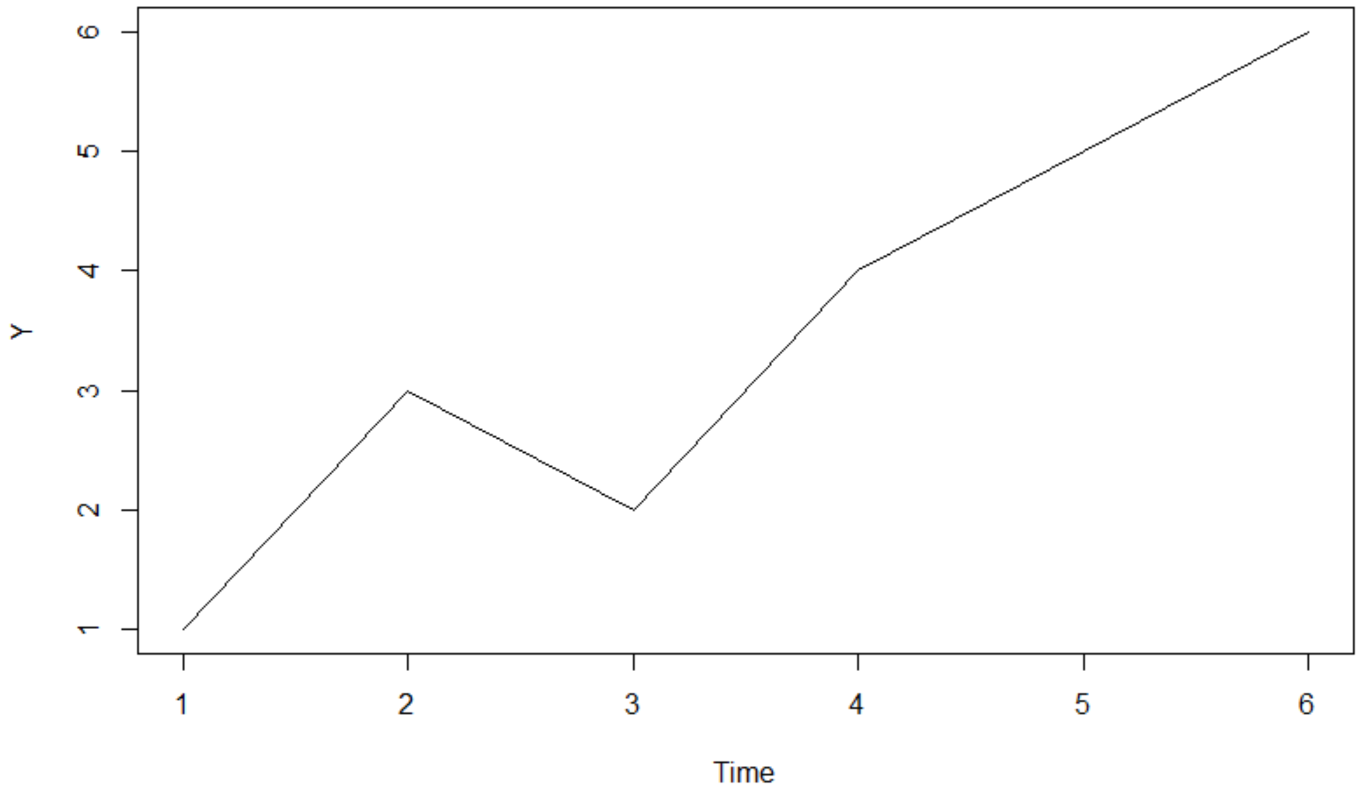
Two Special ESM Models



Weights applied to past values to predict Y_9

- Sample mean: equal weights
- Random walk: all weight to the most recent observation

Example: Predicting Y_7

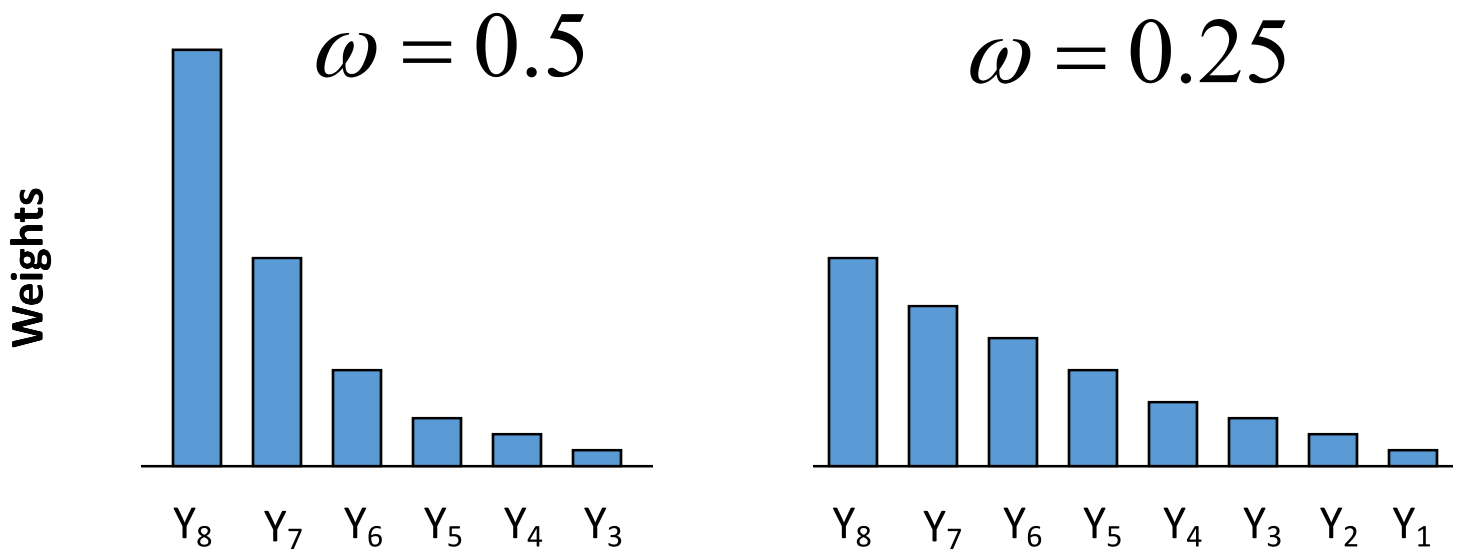


Prediction by sample mean:
Prediction by random walk:

Simple Exponential Smoothing

$$
\begin{aligned}
\hat{Y}_{t+1} &= \omega Y_t + (1-\omega)\hat{Y}_t \\
&= \omega Y_t + (1-\omega)[\omega Y_{t-1} + (1-\omega)\hat{Y}_{t-1}] \\
&= \omega Y_t + \omega(1-\omega)Y_{t-1} + (1-\omega)^2 \hat{Y}_{t-1} \\
&= \omega Y_t + \omega(1-\omega)Y_{t-1} + (1-\omega)^2[\omega Y_{t-2} + (1-\omega)\hat{Y}_{t-2}] \\
&= \omega Y_t + \omega(1-\omega)Y_{t-1} + \omega(1-\omega)^2 Y_{t-2} + \omega(1-\omega)^3 Y_{t-3} + \cdots
\end{aligned}
$$

Simple Exponential Smoothing



Weights applied to past values to predict Y_9

The larger the parameter, the more that
the most recent values are emphasized.

What Happens if ω Approaches Zero

$$
\begin{aligned}
\hat{Y}_{t+1} &= \omega Y_t + (1-\omega)\hat{Y}_t \\
&= \omega Y_t + (1-\omega)[\omega Y_{t-1} + (1-\omega)\hat{Y}_{t-1}] \\
&= \omega Y_t + \omega(1-\omega)Y_{t-1} + (1-\omega)^2 \hat{Y}_{t-1} \\
&= \omega Y_t + \omega(1-\omega)Y_{t-1} + (1-\omega)^2[\omega Y_{t-2} + (1-\omega)\hat{Y}_{t-2}] \\
&= \omega Y_t + \omega(1-\omega)Y_{t-1} + \omega(1-\omega)^2 Y_{t-2} + \omega(1-\omega)^3 Y_{t-3} + \cdots
\end{aligned}
$$

Since $1-\omega \approx 1$, the ESM model degenerates to the sample mean model

ESM Parameters

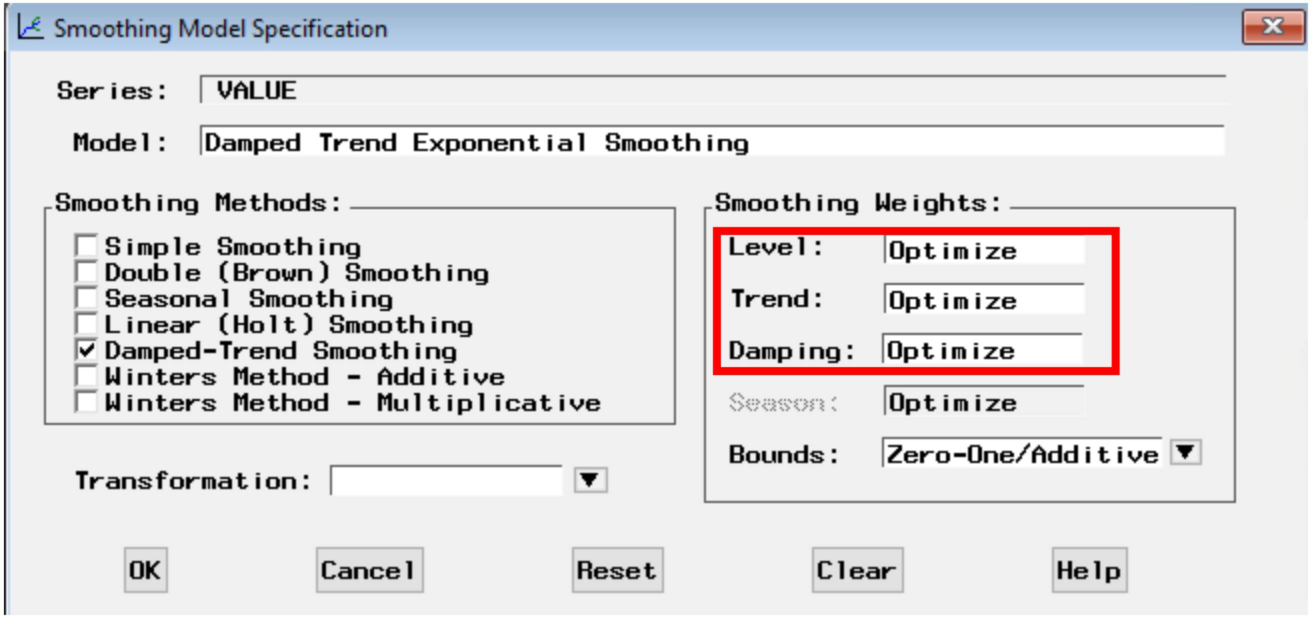
ESM	Parameters	Component
Simple	ω	Level
Double	ω	Level/Trend
Linear (Holt)	ω, γ	Level,Trend
Damped-Trend	ω, γ, ϕ	Level,Trend,Damping

Smoothing Weights

ω Level smoothing weight

γ Trend smoothing weight

ϕ Trend damping weight



Simple Exponential Smoothing Predictions

$\widehat{Y}_1 = Y_0$ (starting value)

$\widehat{Y}_2 = \omega Y_1 + (1-\omega)\widehat{Y}_1$

…

$\widehat{Y}_{t+1} = \omega Y_t + (1-\omega)\widehat{Y}_t$

- The starting value Y_0 is often taken to be the mean of the first *n* observations. SAS TSFS uses *n*=6.
- Does not work well when there is a trend

Double Exponential (Holt) Smoothing

$$L_t = \omega Y_t + (1-\omega)(L_{t-1} + T_{t-1}) \;\; 0 \le \omega \le 1$$

Level equation

$$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1} \;\; 0 \le \gamma \le 1$$

Trend equation

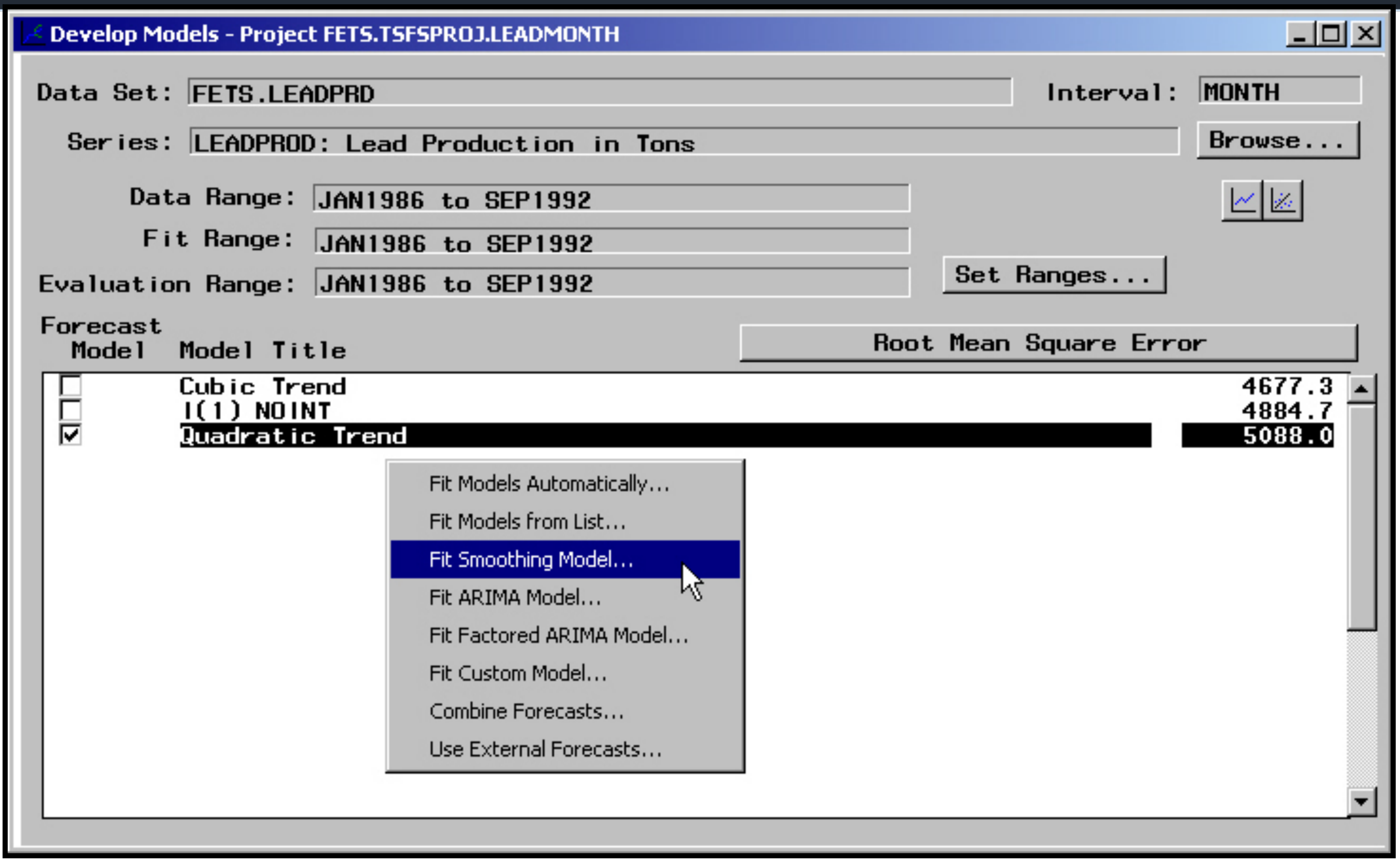
$\hat{Y}_{t+m} = L_t + mT_t$ (m-period-ahead forecast)

Prediction equation

- Smoothing for both level and trend
- TSFS uses Double Exponential Smoothing to refer to a simpler model with only one smoothing parameter (see slide 22)
- Damped-trend smoothing: a third weight on T_{t-1}

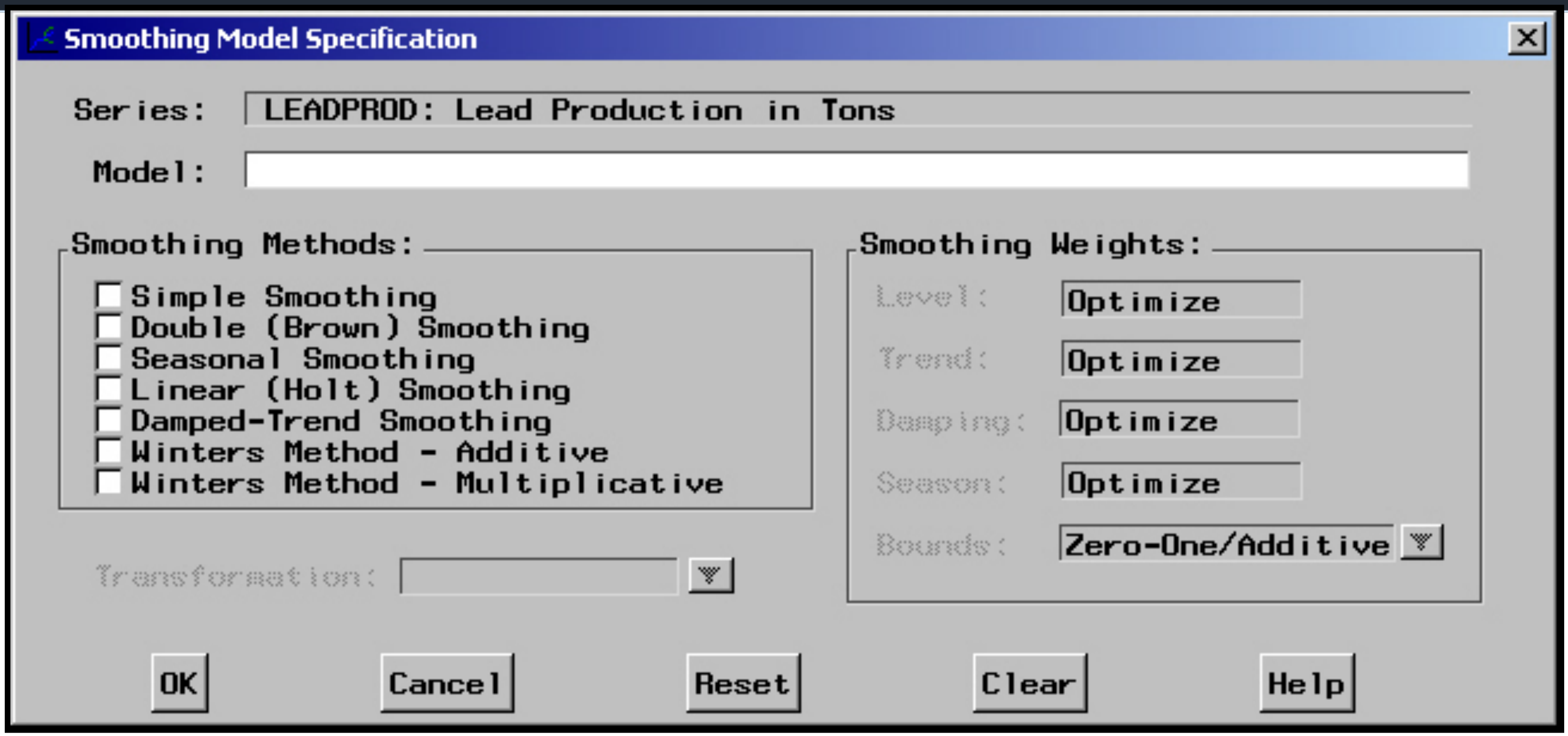
Fore more details
https://onlinecourses.science.psu.edu/stat501/node/363
http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc4.htm

Exponential Smoothing Models in the TSFS



continued...

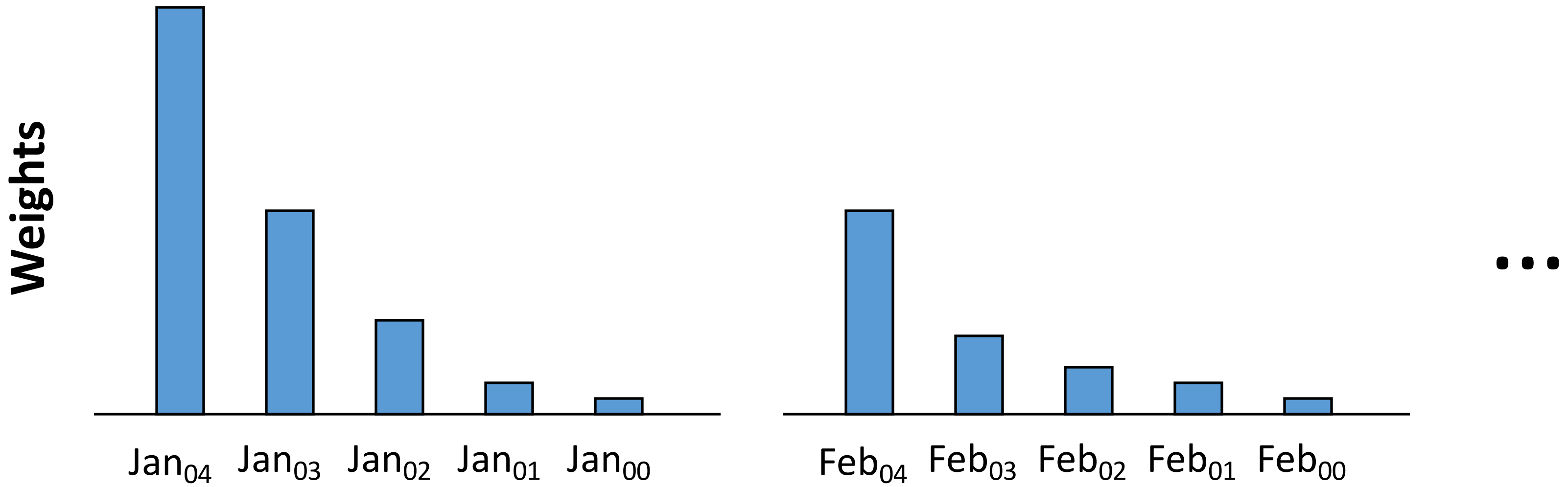
Exponential Smoothing Models in the TSFS



continued...

Seasonal Exponential Smoothing

ESM for Seasonal Data



Weights decay with respect to the seasonal factor.

ESM Parameters (Full)

ESM	Parameters
Simple	ω
Double	ω
Linear (Holt)	ω, γ
Damped-Trend	ω, γ, ϕ
Seasonal	ω, δ
Additive Winters	ω, γ, δ
Multiplicative Winters	ω, γ, δ

Smoothing Weights

ω	Level smoothing weight
γ	Trend smoothing weight
δ	Seasonal smoothing weight
ϕ	Trend damping weight

The choice of Greek letter is arbitrary. The software uses names rather than Greek symbols.

Winters Method — Additive

$$L_t = \omega\left(Y_t - S_{t-p}\right) + (1-\omega)(L_{t-1} + T_{t-1})$$

Level equation

$$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1}$$

Trend equation

$$S_t = \delta(Y_t - L_t) + (1-\delta)S_{t-p}$$

Seasonality equation

$Y_{t+m} = L_t + mT_t + S_t$ (m-period-ahead forecast)

Prediction equation

p is the period of seasonality

Winters Method — Multiplicative

$$L_t = \omega\left(Y_t/S_{t-p}\right) + (1-\omega)(L_{t-1} + T_{t-1})$$

Level equation

$$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1}$$

Trend equation

$$S_t = \delta(Y_t/L_t) + (1-\delta)S_{t-p}$$

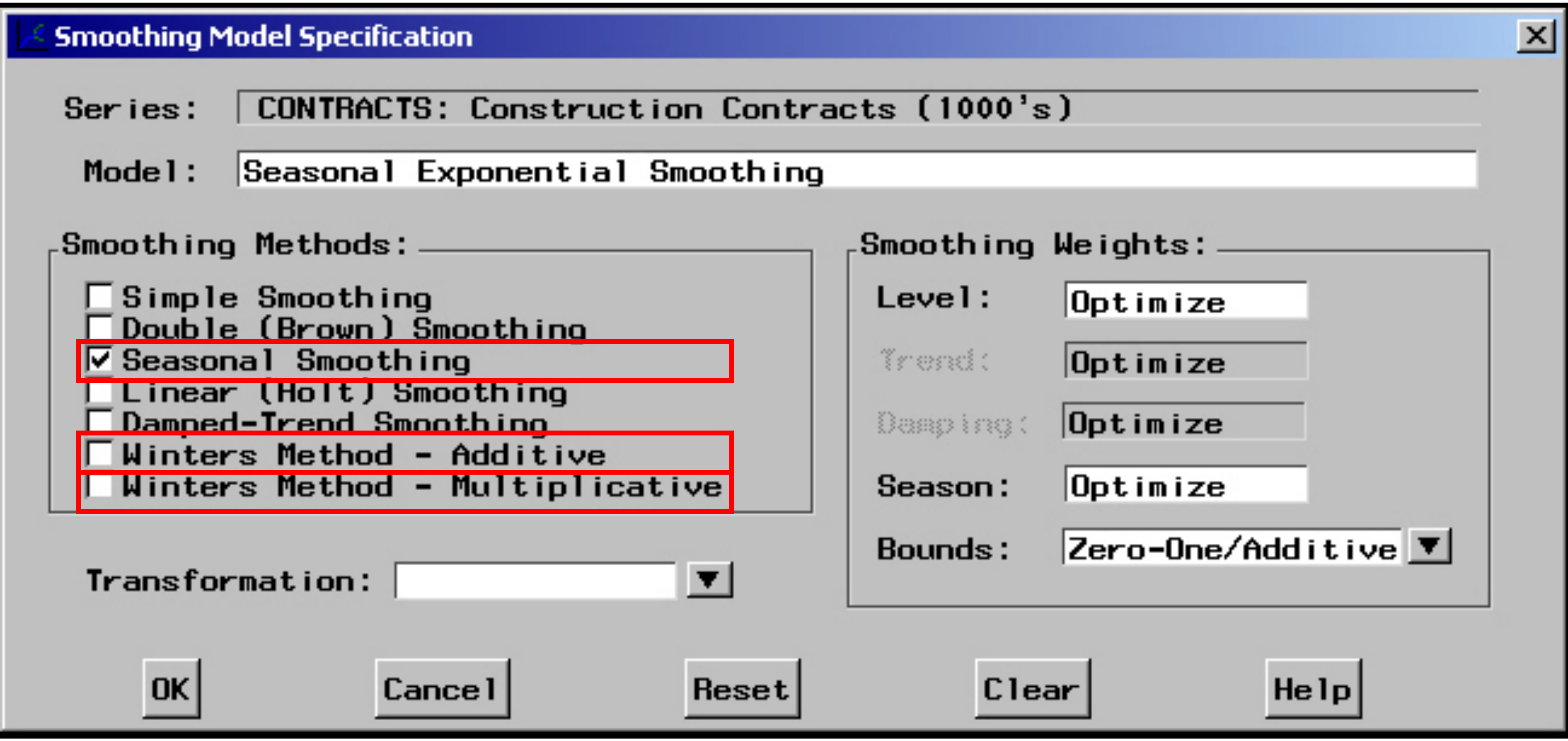
Seasonality equation

$Y_{t+m} = (L_t + mT_t)S_t$ (m-period-ahead forecast)

Prediction equation

p is the period of seasonality

Seasonal Exponential Smoothing Models in the TSFS



continued...

Applications

Fitting ESM for LEADPRD Data

- Models
  - Fit models automatically (use “Diagnose Series” to force SAS to fit all 42 predefined models)
  - Try all non-seasonal ESM models
  - Fit AR(1) and MA(1) models

- Time Series Diagnostics

- Model Residuals Diagnostics

Fitting Seasonal ESM for CONTRACTS Data

- Fit models automatically
- Fit seasonal ESM models
- Setup holdout sample and prediction horizon
- Change evaluation metrics and compare models
- Save predictions

When to Use?

- The prediction of ESM models quickly converges to the mean of time series (plus trend adjustment), not ideal for long-term prediction
  - Simple ESM: short-term prediction
  - Complicated ESM: moderate-term prediction

Readings

- Forecasting Chapter 2
- http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc4.htm
- https://onlinecourses.science.psu.edu/stat501/node/363
- https://people.duke.edu/~rnau/411avg.htm