Time Series and Forecasting

Time Series

* A time series is a sequence of measurements over time, usually obtained at equally spaced intervals
  + Daily
  + Monthly
  + Quarterly
  + Yearly

Time Series Example

Dow Jones Industrial Average

12000

11000

10000

9000

8000

7000

1/3/00 5/3/00 9/3/00 1/3/01 5/3/01 9/3/01 1/3/02 5/3/02 9/3/02 1/3/03 5/3/03 9/3/03

**Date**

**Closing Value**

Components of a Time Series

* Secular Trend
  + Linear
  + Nonlinear
* Cyclical Variation
  + Rises and Falls over periods longer than one year
* Seasonal Variation
  + Patterns of change within a year, typically repeating themselves
* Residual Variation

Components of a Time Series

Yt = Tt + Ct + St + Rt

Time Series with Linear Trend

Yt = a + b t + et

Time Series with Linear Trend

AOL Subscribers

30

25

20

15

10

5

0

2 3

1995

4

1

2

1996

3

4

1 2

1997

3

4

1 2

3

4

1

2

1999

3

4

**Quarter**

1 2

2000

3

1998

Time Series with Linear Trend

**Average Daily Visits in August to Emergency Room**

**at Richmond Memorial Hospital**

140

120

100

80

60

40

20

0

1

2

3

4

5

6

7

8

9

10

**Year**

**Number of Subscribers (millions)**

**A verag e D a ily V isits**

Time Series with Nonlinear Trend

**Imports**

180

160

140

120

100

80

60

40

20

0

1986

1988

1990

1992

1994

1996

1998

**Year**

**Imports (MM)**

Time Series with Nonlinear Trend

* Data that increase by a constant amount at each successive time period show a linear trend.
* Data that increase by increasing amounts at each successive time period show a curvilinear trend.
* Data that increase by an equal percentage at each successive time period can be made linear by applying a logarithmic transformation.

Transformed Time Series

Log Imports

2.5

2.0

1.5

1.0

0.5

0.0

1986

1988

1990

1992

1994

1996

1998

**Year**

**Log(Imports)**

Nonlinear Time Series transformed to a Linear Time Series with a Logarithmic Transformation

log(Yt) = a + b t + et

Time Series with both Trend and

Seasonal Pattern

**Quarterly Power Loads**

200

175

150

125

100

75

50

1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4

1988

1989

1990

1991

1992

1993

1994 1995

1996

1997

1998

1999

**Year and Quarter**

**Power Load**

Model Building

* For the Power Load data
  + What kind of trend are we seeing?
    - Linear
    - Logarithmic
    - Polynomial
  + How can we smooth the data?
  + How do we model the distinct seasonal pattern?

Power Load Data with Linear Trend

**Quarterly Power Loads**

200

y = 1.624t + 77.906

175

R2 = 0.783

Linear Trend Line

150

125

100

75

50

1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4

1988

1989

1990

1991

1992

1993

1994 1995

1996

1997

1998

1999

**Year and Quarter**

**Power Load**

Modeling a Nonlinear Trend

* If the time series appears to be changing at an increasing rate over time, a logarithmic model in Y may work:

ln(Yt) = a + b t + et

or

Yt = exp{a + b t + et }

* In Excel, this is called an exponential model

Power Load Data

with Exponential Trend

**Quarterly Power Loads**

200

175

y = 79.489e0.0149x

R2 = 0.758

Logarithmic (in y) Trend Line

150

125

100

75

Ln(y) = 4.376 + 0.0149t

R2 = 0.758

50

1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4

1988

1989

1990

1991

1992

1993

1994 1995

1996

1997

1998

1999

**Year and Quarter**

**Power Load**

Modeling a Nonlinear Trend

* If the time series appears to be changing at a decreasing rate over time, a logarithmic model in t may work:

Yt = a + b ln(t) + et

* In Excel, this is called a logarithmic model

Power Load Data

with Logarithmic Trend

**Quarterly Power Loads**

200

175

y = 25.564Ln(t) + 42.772

R2 = 0.7778

Logarithmic (in t) Trend Line

150

125

100

75

50

1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4

1988

1989

1990

1991

1992

1993

1994 1995

1996

1997

1998

1999

**Year and Quarter**

**Power Load**

Modeling a Nonlinear Trend

* General curvilinear trends can often be model with a polynomial:
  + Linear (first order)

Yt = a + b t + et

* + Quadratic (second order)

Yt = a + b1 t + b2 t2 + et

* + Cubic (third order)

Yt = a + b1 t + b2 t2 + b3 t3 + et

Power Load Data

modeled with

Second Degree Polynomial Trend

**Quarterly Power Loads**

200

175

y = -0.0335t2 + 3.266t + 64.222

R2 = 0.8341

Second Order Polynomial Trend Line

150

125

100

75

50

1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4

1988

1989

1990

1991

1992

1993

1994 1995

1996

1997

1998

1999

**Year and Quarter**

**Power Load**

Moving Average

* Another way to examine trends in time series is to compute an average of the last m consecutive observations
* A 4-point moving average would be:

y

MA(4)

= (yt + yt-1 + yt-2 + yt-3 )

4

Power Load Data

with 4-point Moving Average

**Quarterly Power Loads**

200

175

150

125

100

75

50

1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4

1988

1989

1990

1991

1992

1993

1994 1995

1996

1997

1998

1999

**Year and Quarter**

**Power Load**

Moving Average

* In contrast to modeling in terms of a mathematical equation, the moving average merely smooths the fluctuations in the data.
* A moving average works well when the data have
  + a fairly linear trend
  + a definite rhythmic pattern of fluctuations

Power Load Data

with 8-point Moving Average

**Quarterly Power Loads**

200

175

150

125

100

75

50

1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4

1988

1989

1990

1991

1992

1993

1994 1995

1996

1997

1998

1999

**Year and Quarter**

**Power Load**

Exponential Smoothing

* An exponential moving average is a weighted average that assigns positive weights to the current value and to past values of the time series.
* It gives greater weight to more recent values, and the weights decrease exponentially as the series goes farther back in time.

Exponentially Weighted Moving Average

S1 = Y1

St = wYt + (1- w)St-1

= wYt + w(1- w)Yt-1 + w(1- w)2 Yt-2 +…

Exponentially Weighted Moving Average

Let w=0.5

S1 = Y1

S2 = 0.5Y2 + (1- 0.5)S1  0.5Y2 + 0.5Y1

S3 = 0.5Y3 + (1- 0.5)S2  0.5Y3 + 0.25Y2 + 0.25Y1

S4 = 0.5Y4 + (1- 0.5)S3  0.5Y4 + 0.25Y3 + 0.125Y2 + 0.125Y1

Exponential Weights

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **w** | **w\*(1-w)** | **w\*(1-w)2** | **w\*(1-w)3** | **w\*(1-w)4** |
| 0.01 | 0.0099 | 0.0098 | 0.0097 | 0.0096 |
| 0.05 | 0.0475 | 0.0451 | 0.0429 | 0.0407 |
| 0.1 | 0.0900 | 0.0810 | 0.0729 | 0.0656 |
| 0.2 | 0.1600 | 0.1280 | 0.1024 | 0.0819 |
| 0.3 | 0.2100 | 0.1470 | 0.1029 | 0.0720 |
| 0.4 | 0.2400 | 0.1440 | 0.0864 | 0.0518 |
| 0.5 | 0.2500 | 0.1250 | 0.0625 | 0.0313 |
| 0.6 | 0.2400 | 0.0960 | 0.0384 | 0.0154 |
| 0.7 | 0.2100 | 0.0630 | 0.0189 | 0.0057 |
| 0.8 | 0.1600 | 0.0320 | 0.0064 | 0.0013 |
| 0.9 | 0.0900 | 0.0090 | 0.0009 | 0.0001 |
| 0.95 | 0.0475 | 0.0024 | 0.0001 | 0.0000 |
| 0.99 | 0.0099 | 0.0001 | 0.0000 | 0.0000 |

Exponential Smoothing

* The choice of w affects the smoothness of Et.
  + The smaller the value of w, the smoother the plot of Et.
  + Choosing w close to 1 yields a series much like the original series.

Power Load Data

with Exponentially Weighted Moving Average (w=.34)

**Quarterly Power Loads**

200

175

150

125

100

75

50

1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4

1988

1989

1990

1991

1992

1993

1994 1995

1996

1997

1998

1999

**Year and Quarter**

**Power Load**

Forecasting with Exponential Smoothing

* The predicted value of the next observation is the exponentially weighted average corresponding to the current observation.

yˆ t+1 = St

Assessing the Accuracy

of the Forecast

* Accuracy is typically assessed using either the Mean Squared Error or the Mean Absolute Deviation

MSE = t=1

 

n

y - y

t t

ˆ



2

n

MAD = t=1

 yt - yˆ t

n

n

Assessing the Accuracy of the Forecast

* It is usually desirable to choose the weight w to minimize MSE or MAD.
* For the Power Load Data, the choice of w = .34 was based upon the minimization of MSE.

Power Load Data with Forecast for

2000 using Exponentially Weighted Moving Average (w=.34)

**Quarterly Power Loads**

200

175

150

125

100

75

50

1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4

1988

1989

1990

1991

1992

1993

1994

1995

1996

1997 1998

1999

2000

**Year and Quarter**

**Power Load**

Exponential Smoothing

* One Parameter (w)
  + Forecasted Values beyond the range of the data, into the future, remain the same.
* Two Parameter
  + Adds a parameter (v) that accounts for trend in the data.

2 Parameter Exponential Smoothing

St = wYt + (1- w)St-1 + Tt-1

Tt = vSt - St-1 + (1- v)Tt-1

The forecasted value of y is

yˆ t+1 = St + Tt

2 Parameter Exponential Smoothing

* The value of St is a weighted average of the current observation and the previous forecast value.
* The value of Tt is a weighted average of the change in St and the previous estimate of the trend parameter.

Power Load Data with Forecast

for 2000 using 2 Parameter Exponential Smoothing with w=.34 and v=.08

**Quarterly Power Loads**

200

175

150

125

100

75

50

1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4

1988

1989

1990

1991

1992

1993

1994

1995

1996

1997 1998

1999

2000

**Year and Quarter**

**Power Load**

Exponential Smoothing

* One Parameter (w)
  + Determines location.
* Two Parameter
  + Adds a parameter (v) that accounts for trend in the data.
* Three Parameter
  + Adds a parameter (c) that accounts for seasonality.

3 Parameter

Exponential Smoothing

S = w + (1- w) S

Y

t

t

I



t-1

+ T

t-1



t-p

Tt = vSt - St-1 + (1- v)Tt-1

I = c Yt + (1- c)I

p

S

t-p

n

The forecasted value of y is

yˆ t+1 = St + Tt + It

3 Parameter Exponential Smoothing

* The value of It represents a seasonal index at point p in the season.

Power Load Data with Forecast

for 2000 using 3 Parameter Exponential Smoothing with w=.34, v=.08, c=.15

**Quarterly Power Loads**

200

175

150

125

100

75

50

1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4

1988

1989

1990

1991

1992

1993

1994

1995

1996

1997 1998

1999

2000

**Year and Quarter**

**Power Load**

Modeling Seasonality

* Seasonality can also be modeled using dummy (or indicator) variables in a regression model.

Power Load Data modeled

for both Trend and Seasonality

**Quarterly Power Loads**

200

175

y = -0.0335t2 + 3.278t + 13.66Q1 - 3.8Q1 + 18.4Q3 +56.86 R2 = 0.9655

150

125

100

75

50

1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4

1988

1989

1990

1991

1992 1993

1994 1995

1996

1997

1998

1999

**Year and Quarter**

Pow er Load

Predicted

**Power Load**

Modeling Seasonality

* For the Power Load data, both trend and seasonality can be modeled as follows: Yt = a + b1 t + b2 t2 + b3 Q1 + b4 Q2 + b5 Q3 + et where

Q = ⎧1

1

⎨

⎩0

Q

2

= ⎧1

⎨

⎩0

Q

3

= ⎧1

⎨

⎩0

if quarter 1

if quarters 2, 3, 4

if quarter 2

if quarters 1, 3, 4

if quarter 3

if quarters 1, 2, 4

Predicted Power Loads

**Predicted Quarterly Power Loads**

200

175

y = -0.0335t2 + 3.278t + 13.66Q1 - 3.8Q1 + 18.4Q3 +56.86 R2 = 0.9655

150

125

100

75

50

1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4

1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999

**Year and Quarter**

Predicted

**Power Load**