Examples for 12/8/2015 – Part 1

A **time series** is a sequence of random variables indexed by time.

A **time series** is a set of measurements of a variable that are ordered through time.

A **time series** is a collection of data obtained by observing a response variable at periodic points in time.

Additive model: $$Y_t = T_t + C_t + S_t + R_t$$

T_t – trend component

C_t – cyclical component

S_t – seasonal component

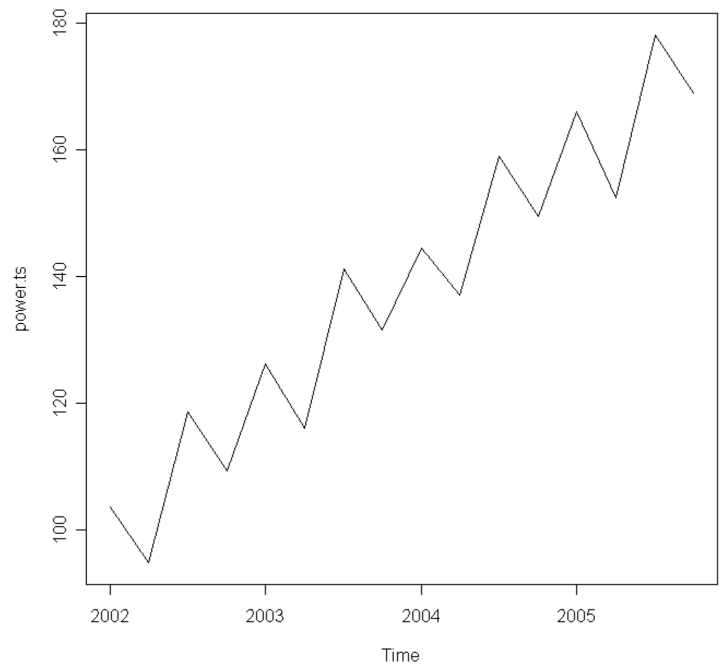
R_t – random (irregular, residual) compoment

Multiplicative model: $$Y_t = T_t \times C_t \times S_t \times R_t$$

Example: Consider the 2002–2005 quarterly power loads for a utility company located in southern Illinois:

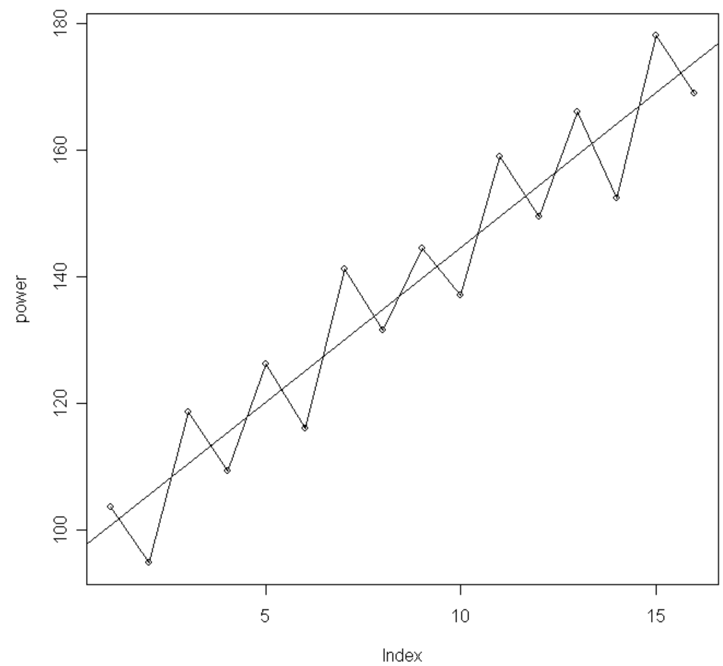
Year	Quarter	Time t	Power Load Y_t, megawatts
2002	I	1	103.5
	II	2	94.7
	III	3	118.6
	IV	4	109.3
2003	I	5	126.1
	II	6	116.0
	III	7	141.2
	IV	8	131.6
2004	I	9	144.5
	II	10	137.1
	III	11	159.0
	IV	12	149.5
2005	I	13	166.1
	II	14	152.5
	III	15	178.2
	IV	16	169.0

```
> power = scan(" ... /power.dat")
Read 16 items
> power
 [1] 103.5  94.7 118.6 109.3 126.1 116.0 141.2 131.6 144.5 137.1 159.0 149.5
[13] 166.1 152.5 178.2 169.0
> power.ts = ts(power, start=2002, frequency=4)
> power.ts
      Qtr1  Qtr2  Qtr3  Qtr4
2002 103.5  94.7 118.6 109.3
2003 126.1 116.0 141.2 131.6
2004 144.5 137.1 159.0 149.5
2005 166.1 152.5 178.2 169.0
> plot(power.ts)
```




```
> t = 1:16
> t
 [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16
> fit = lm(power ~ t)
> summary(fit)                    # estimating trend component

Call:
lm(formula = power ~ t)

Residuals:
     Min       1Q   Median       3Q      Max
-11.7522  -6.3585  -0.1592   7.1202  11.2426

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  95.6625     4.3573   21.95 3.03e-12 ***
t             4.8993     0.4506   10.87 3.28e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.309 on 14 degrees of freedom
Multiple R-squared: 0.8941,     Adjusted R-squared: 0.8865
F-statistic: 118.2 on 1 and 14 DF,  p-value: 3.284e-08
```

```
> plot(power)
> lines(power)
> abline(fit$coeff)
```



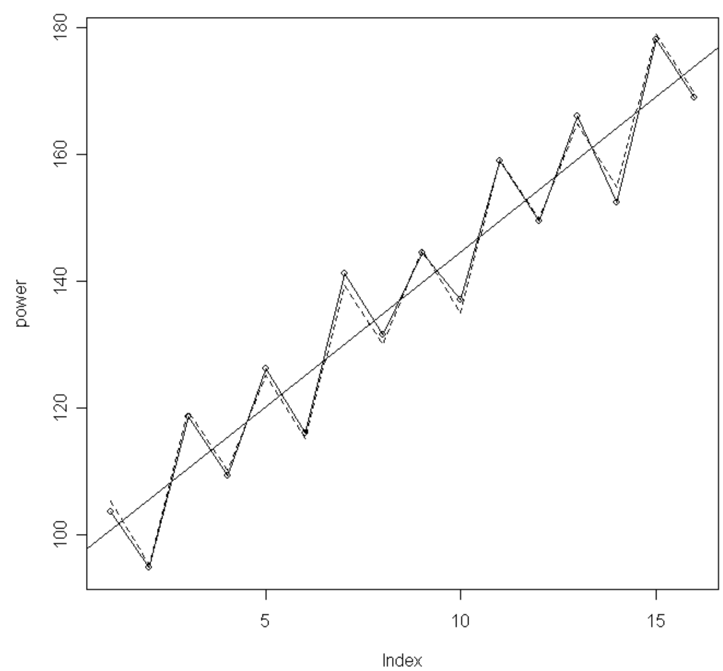
Forecasts:

1st quarter of 2006	$t = 17$	$95.6625 + 4.8993 \times 17 = 178.9506$
2nd quarter of 2006	$t = 18$	$95.6625 + 4.8993 \times 18 = 183.8499$

and so on.

```
> s2 = c(rep(c(0,1,0,0),4))
> s2
 [1] 0 1 0 0 0 1 0 0 0 1 0 0 0 1 0 0
> s3 = c(rep(c(0,0,1,0),4))
> s3
 [1] 0 0 1 0 0 0 1 0 0 0 1 0 0 0 1 0
> s4 = (rep(c(0,0,0,1),4))
> s4
 [1] 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 1
> fit2 = lm(power ~ t + s2 + s3 + s4)        # estimating trend
> summary(fit2)                              # and seasonal components

Call:
lm(formula = power ~ t + s2 + s3 + s4)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.3613 -0.7819 -0.3787  1.0500  2.0962 
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 100.29937    0.97297 103.086  < 2e-16 ***
t             4.96437    0.08566  57.951 4.99e-15 ***
s2          -14.93937    1.08696 -13.744 2.85e-08 ***
s3            4.27125    1.09704   3.893  0.00250 ** 
s4          -10.09312    1.11364  -9.063 1.96e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.532 on 11 degrees of freedom
Multiple R-squared: 0.9972,     Adjusted R-squared: 0.9961 
F-statistic:   969 on 4 and 11 DF,  p-value: 6.262e-14

> lines(fit2$fitted.values,lty=2)
```



Forecasts:

1st quarter of 2006 $t = 17$

$$100.29937 + 4.96437 \times 17 = 184.69366$$

2nd quarter of 2006 $t = 18$

$$100.29937 + 4.96437 \times 18 - 14.93937 = 174.71866$$

3rd quarter of 2006 $t = 19$

$$100.29937 + 4.96437 \times 19 + 4.27125 = 198.89365$$

and so on.

Moving Average:

Year	Quarter	Time, t	Yt	Mt	Ratio	Seasonal index
2002	I	1	103.5			1.058413
	II	2	94.7			0.94806
	III	3	118.6	106.525	1.113354	1.08966
	IV	4	109.3	112.175	0.97437	0.979678
2003	I	5	126.1	117.5	1.073191	
	II	6	116	123.15	0.941941	
	III	7	141.2	128.725	1.096912	
	IV	8	131.6	133.325	0.987062	
2004	I	9	144.5	138.6	1.042569	
	II	10	137.1	143.05	0.958406	
	III	11	159	147.525	1.077783	
	IV	12	149.5	152.925	0.977603	
2005	I	13	166.1	156.775	1.05948	
	II	14	152.5	161.575	0.943834	
	III	15	178.2	166.45	1.070592	
	IV	16	169			

M_t – 4-point moving average

Ratio $= y_t / M_t$

Seasonal Index = average of the ratio for a particular quarter.

Forecasts:

$166.45 - 161.575 = 4.875$, $161.575 - 156.775 = 4.8$. Slope ≈ 4.85.

1st quarter of 2006 $(166.45 + 4.85 \times 2) \times 1.058413 = 186.44$

2nd quarter of 2006 $(166.45 + 4.85 \times 3) \times 0.94806 = 171.60$

and so on.

Moving averages are not restricted to 4 points. For example, you may wish to calculate a 7-point moving average for daily data, a 12-point moving average for monthly data, or a 5-point moving average for yearly data. Although the choice of the number of points is arbitrary, you should search for the number of points that yields a smooth series, but not so large that many points at the end of the series are “lost.”

Exponential Smoothing:

Like the moving average method, exponential smoothing tends to deemphasize most of the random (residual) effects.

exponential smoothing constant

$0 < w < 1$

$E_1 = y_1$

$E_2 = w\, y_2 + (1 - w)\, E_1$

$E_3 = w\, y_3 + (1 - w)\, E_2$

…

$E_t = w\, y_t + (1 - w)\, E_{t-1}$

Year	Quarter	Time, t	Yt	Et	w
2002	I	1	103.5	103.5	0.7
	II	2	94.7	97.34	
	III	3	118.6	112.222	
	IV	4	109.3	110.1766	
2003	I	5	126.1	121.323	
	II	6	116	117.5969	
	III	7	141.2	134.1191	
	IV	8	131.6	132.3557	
2004	I	9	144.5	140.8567	
	II	10	137.1	138.227	
	III	11	159	152.7681	
	IV	12	149.5	150.4804	
2005	I	13	166.1	161.4141	
	II	14	152.5	155.1742	
	III	15	178.2	171.2923	
	IV	16	169	169.6877	

Forecasts:

The forecasts for all future periods will be the same:

$F_t = w\, y_N + (1 - w)\, E_N = 0.7 \times 169 + 0.3 \times 169.6877 = 169.2.$

This points out a disadvantage of the exponential smoothing forecasting technique. Since the exponentially smoothed forecast is constant for all future values, any changes in trend and/or seasonality are not taken into account. Therefore, exponentially smoothed forecasts are appropriate only when the trend and seasonal components of the time series are relatively insignificant.