Forecasting Level Time Series

Models

Previously, we looked at **common patterns** in time series

These patterns arose from **systematic components** of time series **influenced by relevant variables**

Examination of typical time series also revealed **random deviations** from expected patterns.

The forecaster's task is to produce a model that can incorporate the **relevant** systematic components while accounting for random variations

Generating Process Model

Although it is tempting to try and produce a model to explain or fit the observed time series sample, forecasting requires the model to accurately predict another sample; **the future**

Thus it is worthwhile to try to find a **general** model that will explain/predict **both past and future samples**

It is worthwhile to try and model the **general process** which generates observations ie **the population model** as opposed to the sample model for the observed time series sample

Generating Process (cont.)

A basic representation of the generating process is;

$$Y_t = f(\text{ systematic, random component})$$

$$Y_t = \mu_t + \varepsilon_t$$

The modeller needs to determine the **relevant functional form** for $\mu_{t.}$

This will depend on the **type of patterns** observed in the time series and the **type of model (time series/causal)**

Random Component

The random component accounts for **unobservable variables, measurement errors and other unpredictable effects**

The random error at **each point in time** is assumed to come from a **symmetric distribution with zero mean.**

The variance of the random error is assumed **constant** over time periods

The random error component at one point of time is assumed **uncorrelated** with random error components at other points of time

Checking the Model Specification

The specified model can be **checked against available evidence** (time series sample)

Correctly specified models will explain the **systematic component** of the generating process

Forecasts from these models will **on average** be accurate (assuming zero mean error)

The residual component overall should be similar to the error component (**symmetric** around a **zero mean**)

In a correctly specified model, the error/ residual component (**variance**) should be **smaller** than for non-correct specifications.

Choosing the Appropriate Model

In reality the task of choosing the correct model specification is not easy.

There is limited information and there may be **competing specifications** that can **theoretically** explain patterns equally well.

Correct specification may also involve **specific parameters** (eg. specfic slope, intercept in a line)

Typically, forecasters will choose the model specification that provides **unbiased forecasts, and reduces error variance.**

Evaluating Forecast Models

To evaluate competing forecast models **(unbiased, min variance)** we use **forecast errors** for the time series **sample** (or part thereof)

Sample forecast error = residual= $e_t = Y_t - F_t$

where F_t = Forecast (estimated model), Y_t = The Actual time series

The forecast models should be evaluated over as **many sample values** as practical.

Initial Check for Correct Model

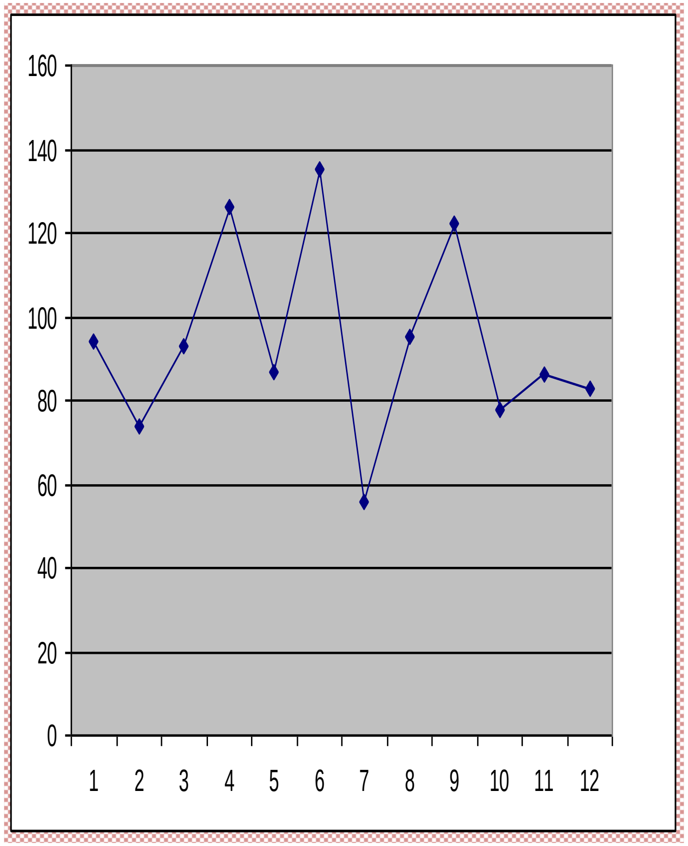
In a correctly specified model, forecasts are **unbiased and the error/residual component will be symmetric with zero mean** (error assumption)

Additionally, the **errors/residuals are uncorrelated over time**

Plots of residuals over time should reflect the above - **mean of zero**, **approximately equal numbers of residuals either side of zero** and **no manifest patterns over time**

This can be checked via examination of the **scatterplot of residuals vs time** and the **ACF**

Example



Consider Sales data for 12 periods:

Examination of the data suggests a **constant level with random fluctuation around that level**.

Plausible generating model is

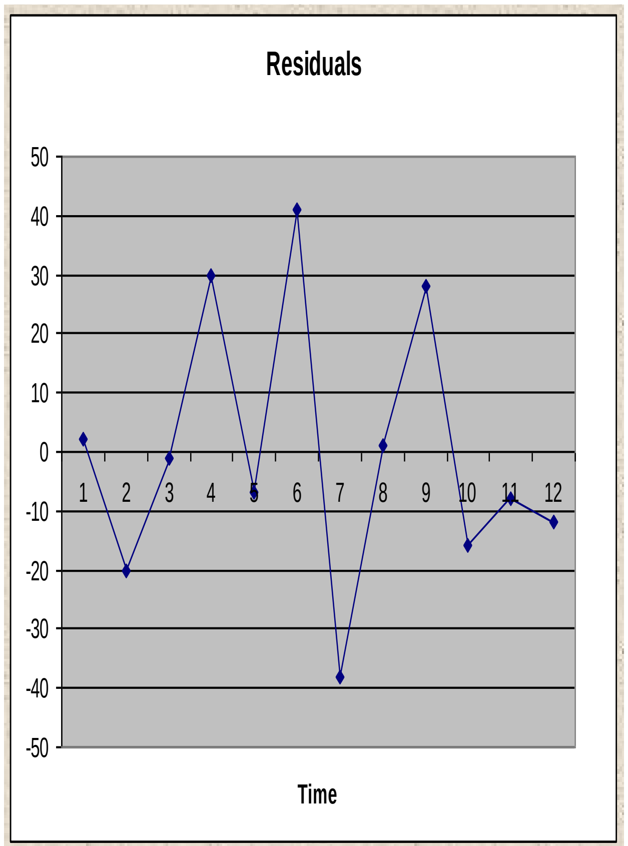
$$\mathbf{Y_t = \mu + \varepsilon_t}$$

Where **μ = constant** (**estimated by average over the time series**)

Forecast = μ (since error cannot be forecast)

MACQUARIE University

Forecasts and Residuals



Time	Data	Forecast	Residual
1	96	94	2
2	74	94	-20
3	93	94	-1
4	124	94	30
5	87	94	-7
6	135	94	41
7	56	94	-38
8	95	94	1
9	122	94	28
10	78	94	-16
11	86	94	-8
12	82	94	-12

Secondary Check on Model

The previous residual checks were undertaken to support the **adequacy** and **“correctness”** of the forecast model specified

In this case, using a constant to forecast the time series appeared to be adequate

However, there is no evidence (as yet) to suggest this model is **minimum variance** (most accurate) among possible competing models

Other models may also be adequate (unbiased forecasts) but produce more accurate forecasts.

Residual/Error Functions

Accuracy check of models is via **error functions**

The error functions are usually based on either the **absolute errors** or the **squared errors** due to the misleading interpretation of the average of errors

Many error functions and no one function is best

A good forecaster will use **several functions as indicators of forecast performance** of models

Sometimes the functions will differ in their choice of best model. The forecaster then must use **other criteria** to decide the preferred model.

MACQUARIE University

Error Functions

Total error $\sum e_t$

Sum of Squared Errors $\sum e_t^2$

Mean Squared Error (MSE) $\frac{\sum e_t^2}{n}$

Root MSE $\sqrt{\left(\frac{\sum e_t^2}{n}\right)}$

More Error Functions

Mean Absolute Error (MAE)

$$\frac{\sum |e_t|}{n}$$

Mean Absolute % Error (MAPE)

$$\frac{\sum |e_t| * 100}{n |X_i|}$$

Using Spreadsheets – Example

Consider a **naïve prediction** for the previous example

A Naïve prediction = **previous periods actual becomes this period's forecast**
Plausible generating model is

$$Y_t = \mu_t + \varepsilon_t$$

Naïve model: $\mu_t = Y_{t-1}$

Forecast = Y_{t-1} (since error cannot be forecast)
Naïve model will only be plausible if the time series is **horizontal** (constant level)

Naïve Forecast Example

Time	Data	Forecast
1	96	
2	74	96
3	93	74
4	124	93
5	87	124
6	135	87
7	56	135
8	95	56
9	122	95
10	78	122
11	86	78
12	82	86

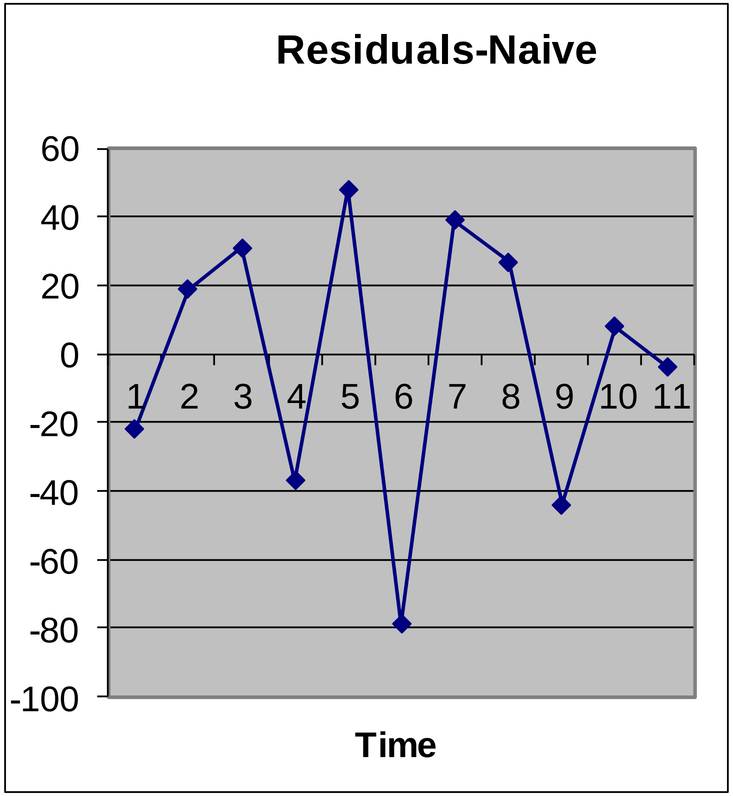
Naïve Forecast – Errors

Time	Data	Forecast	error	abs error	sq error	abs%error
1	96					
2	74	96	-22	22	484	0.297297297
3	93	74	19	19	361	0.204301075
4	124	93	31	31	961	0.25
5	87	124	-37	37	1369	0.425287356
6	135	87	48	48	2304	0.355555556
7	56	135	-79	79	6241	1.410714286
8	95	56	39	39	1521	0.410526316
9	122	95	27	27	729	0.221311475
10	78	122	-44	44	1936	0.564102564
11	86	78	8	8	64	0.093023256
12	82	86	-4	4	16	0.048780488
13		82		MAE	MSE	MAPE
				32.54545	1453.273	0.389172697

Forecast using Average – Errors

Time	Data	Forecast	error	abs error	sq error	abs%error
1	96	94	2	2	4	0.020833333
2	74	94	-20	20	400	0.27027027
3	93	94	-1	1	1	0.010752688
4	124	94	30	30	900	0.241935484
5	87	94	-7	7	49	0.08045977
6	135	94	41	41	1681	0.303703704
7	56	94	-38	38	1444	0.678571429
8	95	94	1	1	1	0.010526316
9	122	94	28	28	784	0.229508197
10	78	94	-16	16	256	0.205128205
11	86	94	-8	8	64	0.093023256
12	82	94	-12	12	144	0.146341463
				MAE	MSE	MAPE
				18.36364	520.3636	0.206383707

Comparison of Forecast Models



The residuals for the naïve model seem to support an error component that is symmetric, zero mean and constant variance

Both models are adequate

However, the model using the **average** seems to be more accurate since it has lower error statistics **(MAE, MSE, MAPE)** in comparison to the naïve model

Smoothing Methods

Smoothing Methods typically seek to **eliminate randomness from the time series** using some type of averaging procedure

This is done using a **smoothing algorithm or set of equations**

The "smoothing" of the time series **more clearly identifies the underlying systematic components**

The systematic components are then used to **make predictions** of the time series

Smoothing Methods (cont.)

- They are reasonably simple to implement
- They are used for short to medium term forecasting
- They are relatively inexpensive

Can be used:

a. when the forecaster needs to make **many predictions on a routine basis**
b. when the forecaster **does not have the resources** for more complex models
c. when the forecaster has **little expertise** in predictive methods

Smoothing Methods (cont.)

Various smoothing methods matched with particular types of time series;

Moving Average >> Horizontal

Simple Exponential Smoothing (SES) >>> Horizontal

Holts Smoothing >>> Trend

Winters Smoothing >>> Seasonality

Averaging

It would seem logical that an **average of past values of the time series may be a good predictor** of future values of the time series.

However, in many cases this is not necessarily so. It will be an appropriate predictor if the time series is reasonably **horizontal** for its entire span with no systematic changes in level.

If the series has trend, seasonal or cycle components the **average of the entire series** will **not typically provide good prediction**

Averages (cont.)

Further it is somewhat **illogical** that all observations in the time series including values in the distant past should have an **equal weighting in predicting future values**

More recent values are more likely to be **better predictors** and should be given **more weight**

The opposite end of the scale is to use a **naïve** predictor which only uses the most recent value to predict future values

If the time series has a **significant random component** or **systematic components (other than Level)** this will not be a good predictor either

Averages (cont.)

The 2 possibilities - **entire average** and **naïve predictors** represent the two extremes

A predictor that may have advantages over both the above is using an average of **a certain number of the most recent observations**

If a fixed number of observations is used for the predictor, as new more recent observations are added the average **should not remain static but "move"**- adding recent values and dropping distant values from the average.

This predictor is a **Moving Average** predictor

Moving Averages

Moving Average (MA) is a predictor for time series that are predominantly **horizontal** ie. the level of the time series remains similar throughout the entire series. (not considering random fluctuation)

The **degree of MA** is the number of observation of the time series used in the predictor

e.g. A **3 period MA** uses the **average of the last 3 observations** in the series

In this regard the last 3 observations have an **equal weight in the predictor (1/3)** while the observations prior to those have **zero weighting in the predictor.**

Moving Average (cont.)

The degree of MA is **arbitrary** however there are some influencing factors;

The volatility of the time series - the more volatile the series (ie larger random component) the more observations should be included in the MA to smooth the series.

The length of the time series - for short times series a short period MA is advised

The predictive performance - several alternative MA predictors can be tried. The MA which has lower error criteria such as **RMSE, MAE or MAPE** may be preferred.

Generating Process – Moving Average

An MA(q) forecast = **average of the previous q periods is the forecast for period t**

The generating model is

$$Y_t = \mu_t + \varepsilon_t$$

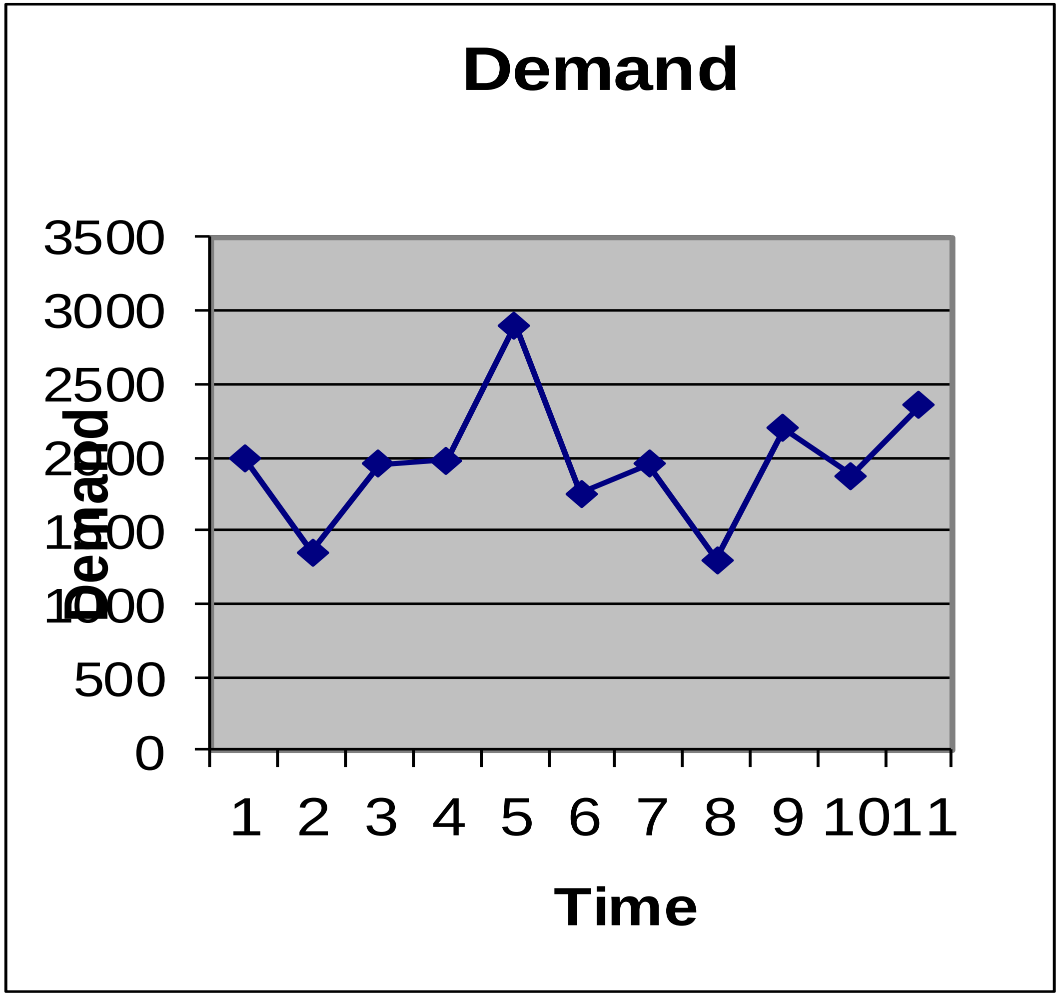
MA(q): $\mu_t = (1/q)\,\{Y_{t-1}+Y_{t-2}+Y_{t-3\ldots}+Y_{t-q}\}$

Forecast = $(1/q)\,\{Y_{t-1}+Y_{t-2}+Y_{t-3\ldots}+Y_{t-q}\}$ (since error cannot be forecast)

MA(q) model will be plausible if the time series is **horizontal** (constant level)

Moving Average Example

Time	Demand
1	2000
2	1350
3	1950
4	1975
5	2900
6	1750
7	1950
8	1300
9	2200
10	1870
11	2350



Simple Exponential Smoothing (SES)

SES is an alternative smoothing method when the time series is predominantly horizontal

The key difference between the two methods is the weights used in the predictor

MA gives every observation in the predictor an equal weighting. Logic would suggest that the most recent observation should have the greatest weight since it will likely be a better predictor of the future values of the time series

SES adopts a weighted averaging procedure with recent observations given relatively more weight.

Generating Process – SES

SES forecast = **weighted average of current actual data and estimated current level or forecast of data**
Plausible generating model is

$$Y_t = \mu_t + \varepsilon_t$$

$$\text{SES}(\alpha): \mu_t = F_t = \alpha^* Y_{t-1} + (1-\alpha) * F_{t-1}$$

α = smoothing parameter (typically between 0 and 1 inclusive)

Forecast (t+1) = $F_{t+1} = \alpha^* Y_t + (1-\alpha) * F_t$ (error cannot be forecast)
SES(α) model will be plausible if the time series is **horizontal** (constant level)

SES (cont.)

As with the choice of the degree of MA the level of a in SES is **arbitrary**. There are some influencing factors

The volatility of the time series- the more volatile the series the lower the value of α needed to smooth the time series. A lower value of α will have the effect of giving observations in the past a greater weighting than if α is higher

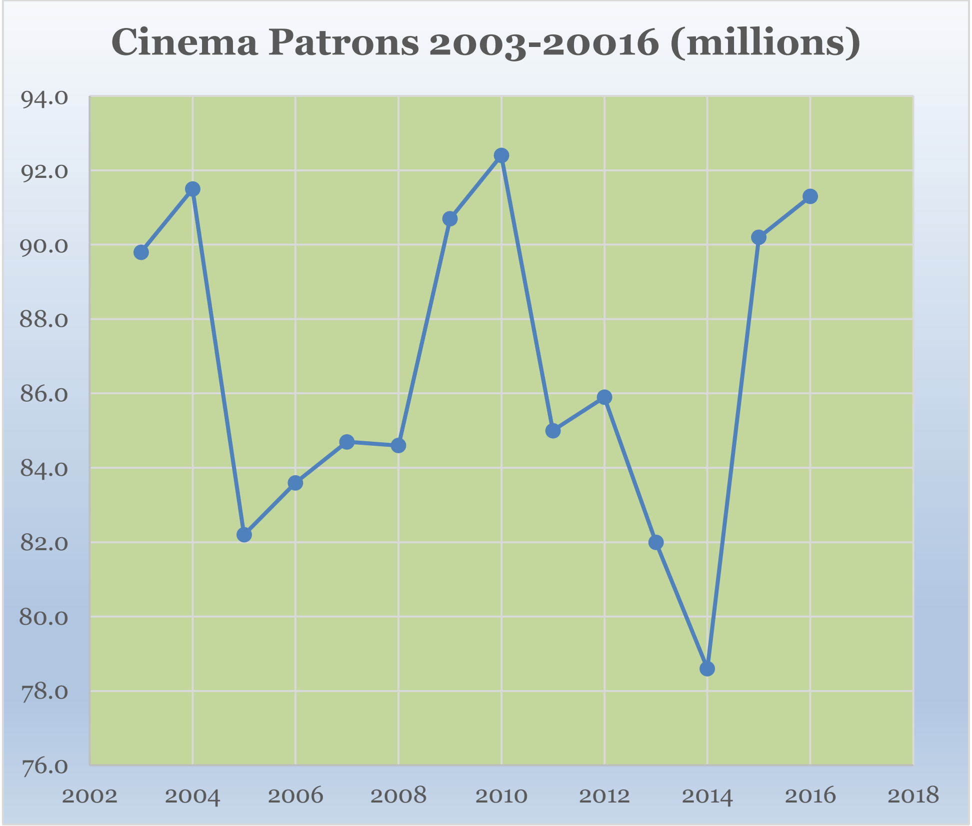
The predictive performance- Different values of the smoothing parameter can be tried on a test set of the time series and the predictive performance compared using error criteria

SOLVER in EXCEL can be used to determine an **"optimum"** α

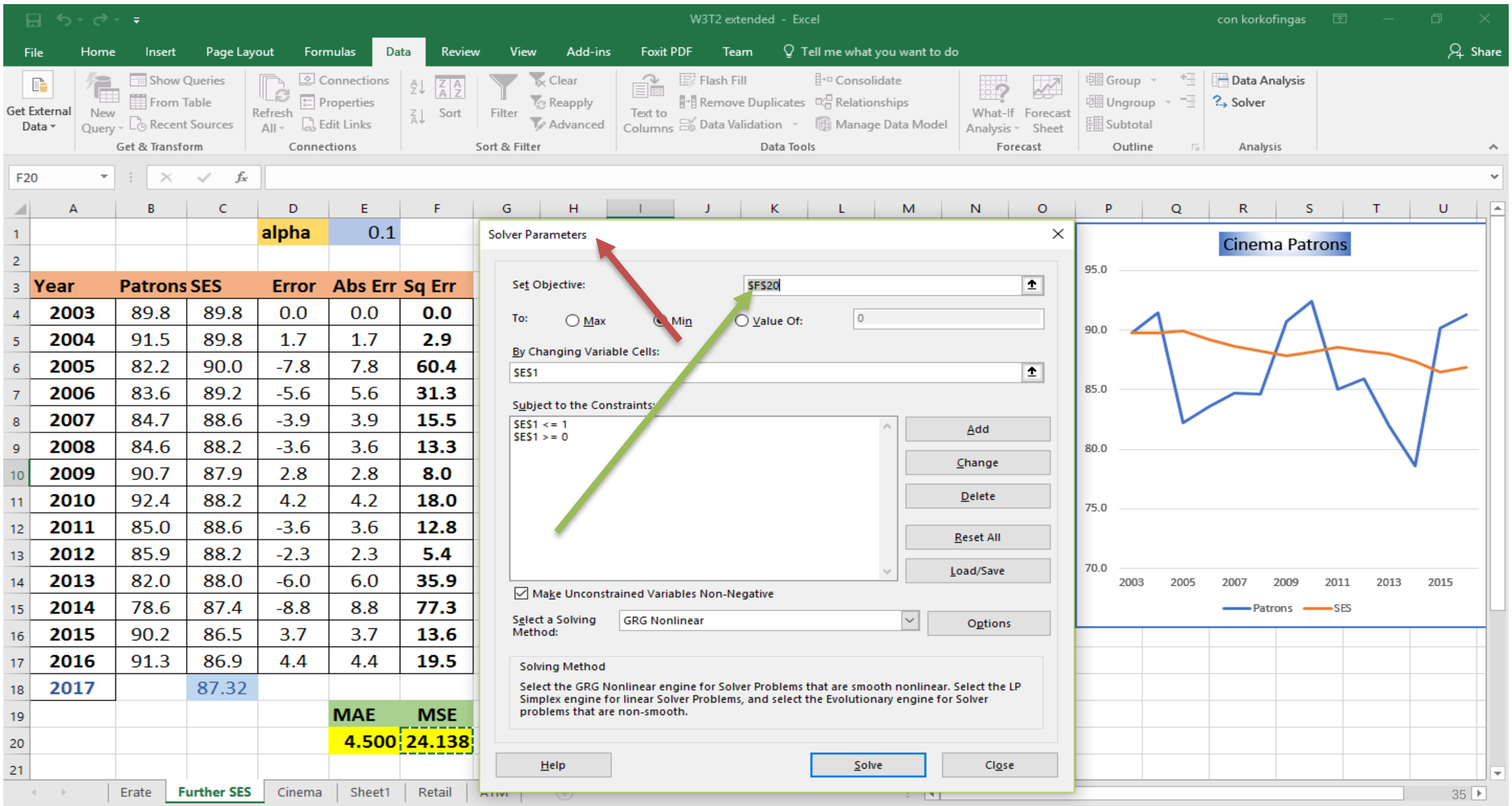
Further SES example

alpha	0.1

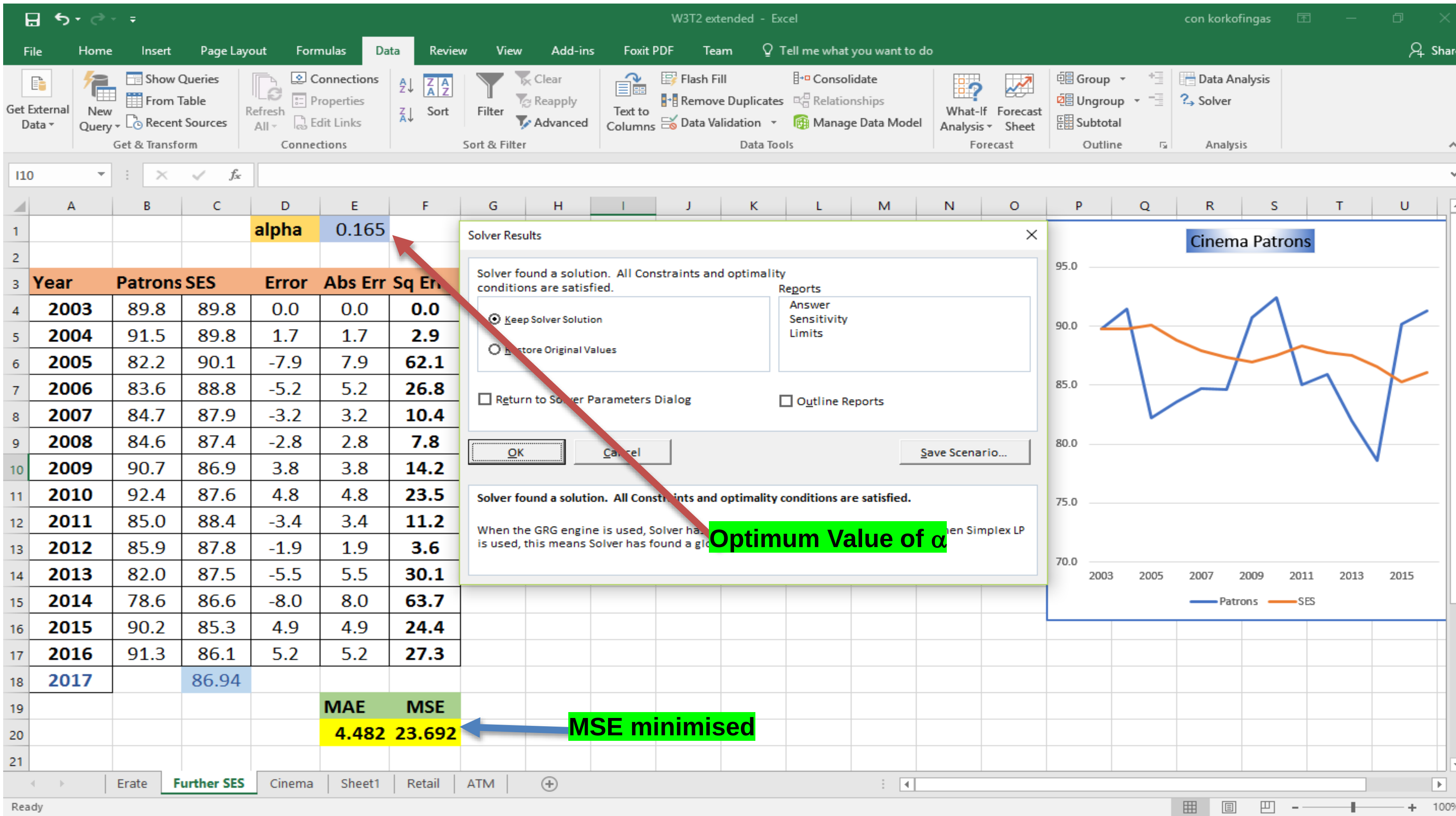
Year	Patrons	SES	Error	Abs Err	Sq Err
2003	**89.8**	89.8	0.0	0.0	**0.0**
2004	**91.5**	89.8	1.7	1.7	**2.9**
2005	**82.2**	90.0	-7.8	7.8	**60.4**
2006	**83.6**	89.2	-5.6	5.6	**31.3**
2007	**84.7**	88.6	-3.9	3.9	**15.5**
2008	**84.6**	88.2	-3.6	3.6	**13.3**
2009	**90.7**	87.9	2.8	2.8	**8.0**
2010	**92.4**	88.2	4.2	4.2	**18.0**
2011	**85.0**	88.6	-3.6	3.6	**12.8**
2012	**85.9**	88.2	-2.3	2.3	**5.4**
2013	**82.0**	88.0	-6.0	6.0	**35.9**
2014	**78.6**	87.4	-8.8	8.8	**77.3**
2015	**90.2**	86.5	3.7	3.7	**13.6**
2016	**91.3**	86.9	4.4	4.4	**19.5**
2017		87.32			
				MAE	**MSE**
				4.500	**24.138**



Using Solver for SES



Solver Solution



Year	Patrons	SES	Error	Abs Err	Sq Err
2003	89.8	89.8	0.0	0.0	0.0
2004	91.5	89.8	1.7	1.7	2.9
2005	82.2	90.1	-7.9	7.9	62.1
2006	83.6	88.8	-5.2	5.2	26.8
2007	84.7	87.9	-3.2	3.2	10.4
2008	84.6	87.4	-2.8	2.8	7.8
2009	90.7	86.9	3.8	3.8	14.2
2010	92.4	87.6	4.8	4.8	23.5
2011	85.0	88.4	-3.4	3.4	11.2
2012	85.9	87.8	-1.9	1.9	3.6
2013	82.0	87.5	-5.5	5.5	30.1
2014	78.6	86.6	-8.0	8.0	63.7
2015	90.2	85.3	4.9	4.9	24.4
2016	91.3	86.1	5.2	5.2	27.3
2017		86.94			