Data Analytics

Introduction to Time Series Analysis & Forecasting

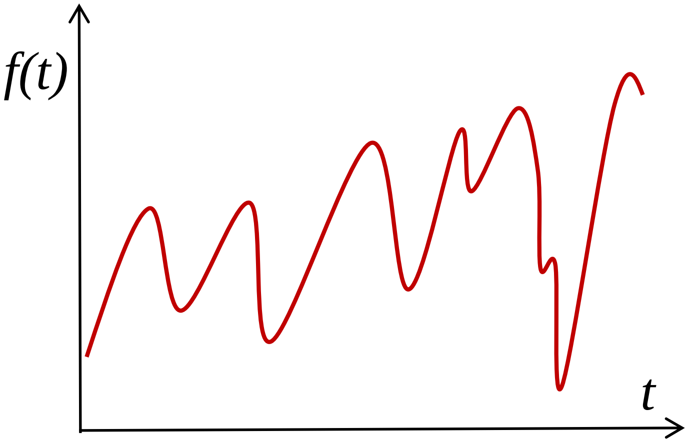
Dr. Rita Chakravarti
Institute of Systems Science
National University of Singapore
Email: rita@nus.edu.sg

Introduction

- What is Time Series
  – It is a chronological sequence of observations on a particular variable
- Examples:
  – Unit sales of a product over time
  – Total dollar sales for a company over time
  – Number of unemployed over time
  – Unemployment rate over time
  – Production of a Product over time
  – Air or water quality over time etc..

What is Time Series Data

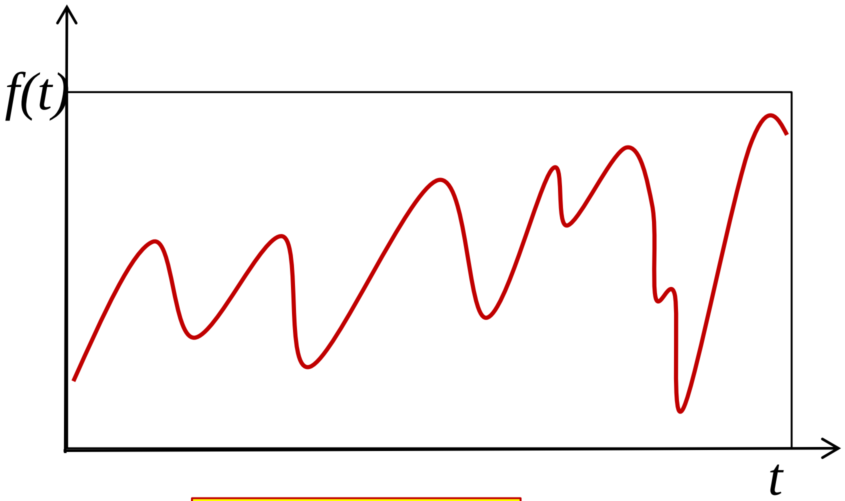
- Important features of time series
  - Usually the observations are taken at regular intervals (e.g.: hours, days, weeks, months, years)
  - The influence among **neighboring observations** is informative for prediction



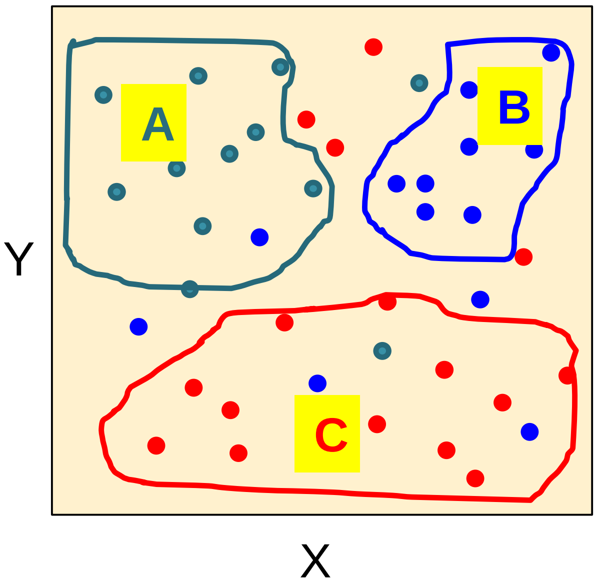
- Some examples of time series
  - the Dow Jones Industrial Average,
  - Gross Domestic Product,
  - unemployment rate, and
  - airline passenger loads
  - … …

Prediction with Time Series

- Time series vs. non time series



Target: f(t)
***No** predictor*



Three categories: A, B, C
***Two** predictors: x, y*

Time Series Analysis

Two main goals :

a) Identifying the nature of phenomenon/**pattern** represented by the data series

b) Forecast/predict the future values

Time Series Patterns & Components

Time series data is assumed to consist of systematic pattern/components and random noise (error). Hence de-noise/noise filtering is required in most time series analysis technique in order to observe the pattern.

The Components of a Time Series are:

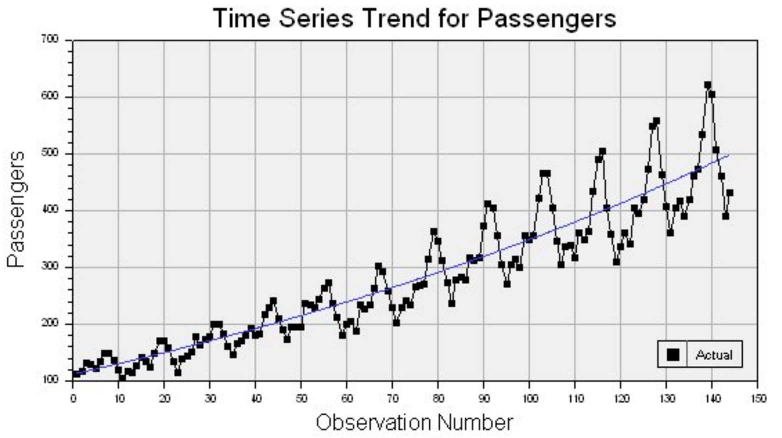
1) Trends
2) Cycle
3) Seasonal variation and
4) Irregular fluctuations

Trend

Trend refers to upward or downward movement that characterizes a time series over a period of time. Thus trend reflects the long-run growth or decline in the time series

Trend movements can represent a variety of factors. For example, long run movements in the sales of a particular industry might be determined by one, some or all of the following factors:

- Technological change in the industry
- Changes in consumer tastes
- Increases in per capita income
- Increases in total population
- Market growth
- Inflation or deflation (price changes)



Cycle

Cycle refers to up and down movements around trend levels. These fluctuations can have a duration of anywhere from two to ten years or even longer measured from peak to peak or trough to trough

One of the common cyclical fluctuations found in time series data is the “business cycle” The business cycle is represented by fluctuations in the time series caused by recurrent periods of prosperity alternating with recession

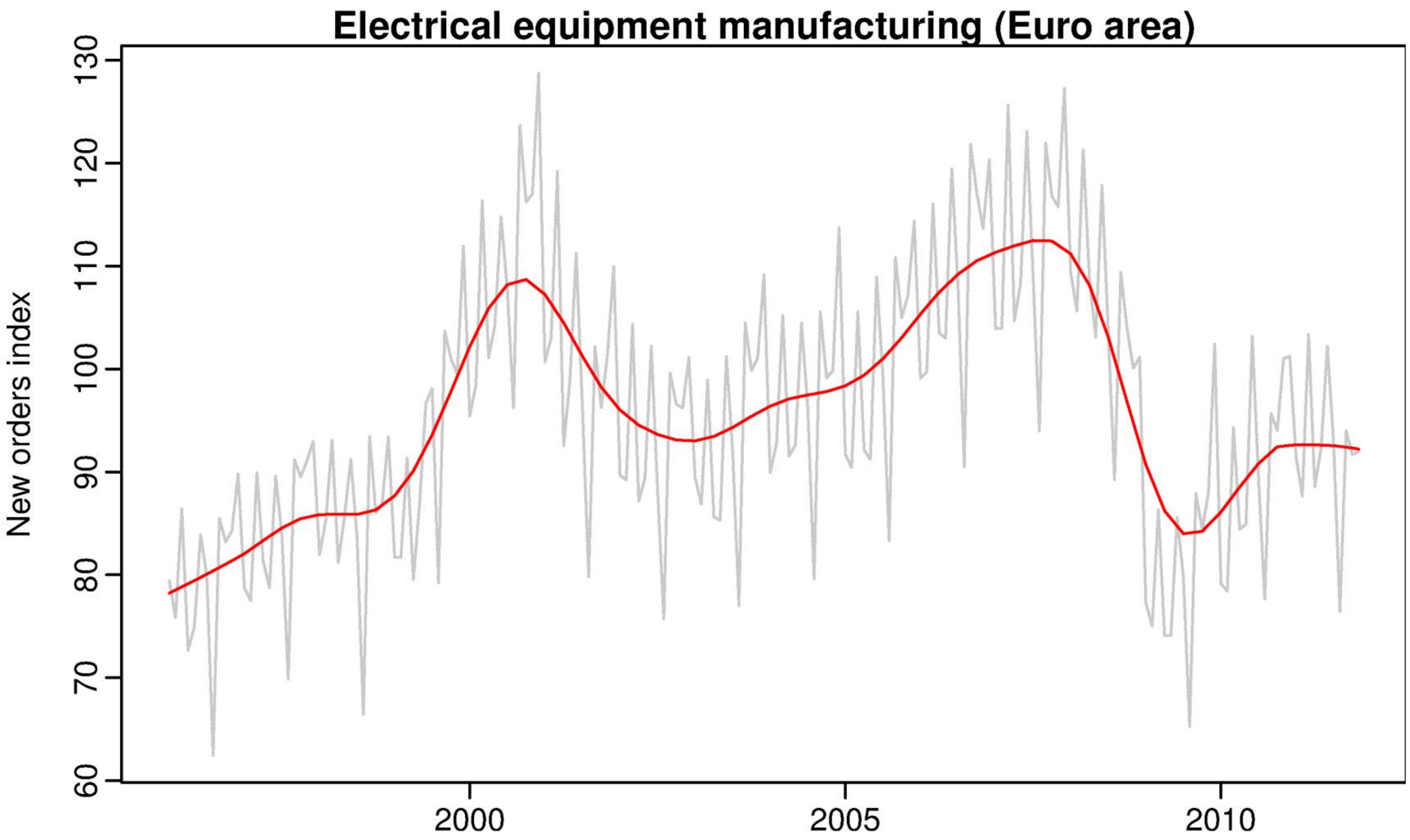
(increase or decrease in economic activity)

However it is not necessary that it always happens due to changes in economic factors e.g.

- Cyclical fluctuations in agricultural yields due to weather cycles
- Cyclical fluctuations in demand of a type of clothing due to fashion change

Because there is no single explanation for cyclical fluctuations the vary greatly in both length and magnitude

Cycle



Electrical equipment manufacturing (Euro area)
New orders index
60
70
80
90
100
110
120
130
2000
2005
2010

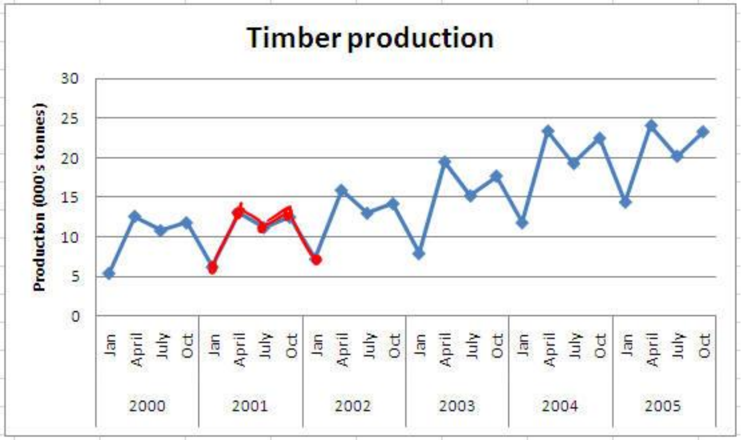
Seasonal Variation

Seasonal variations are periodic patterns in a time series that complete themselves within a calendar year and are then repeated on a yearly basis

Seasonal variations are typically caused by factors such as weather or customs e.g.

- Average monthly temperature
- Monthly sales volume in a departmental stores

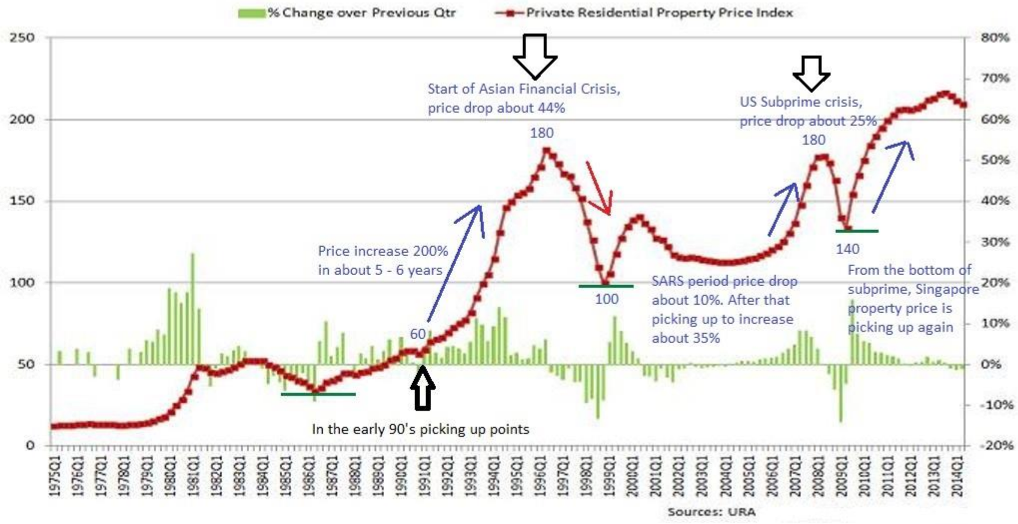
Ordinarily series of monthly or quarterly data are used to examine seasonal variations. Clearly a single yearly observation would not reveal variations that occur during the year



Irregular Fluctuation

Irregular fluctuations are erratic movements in a time series that follow no recognizable regular pattern. Such movements represent what is "left over" in a time series after trend, cycle and seasonal variations have been accounted for.

Many irregular fluctuations in a time series are caused by 'unusual' events that cannot be forecasted – earthquakes, accidents, hurricanes, wars etc.



FORECASTING

Forecasting

- Time series data are often examined in hopes of discovering a historical pattern that can be exploited in the preparation of a forecast which is often used :
  - To make better informed decisions
  - Forecast short term demand, mid-term resource planning, to long term strategic planning

Time Series Forecasting: *where?*

- Time series forecasting is performed in nearly every organization that works with quantifiable data. Examples of use of forecasting for decision support
  - Retail stores: sales
  - Energy companies: reserve, production, demand, prices
  - Education institutes: enrollment
  - Governments: tax receipts and spending
  - Inter. financial organization: inflation and economic activity
  - Transportation companies: future travel
  - Banks: new home purchases
  - Venture capital firms: market potential to evaluate business plan

Time Series Forecast: *as predictive model*

- The goal of building a time series model is the same as the goal for other types of predictive models, e.g.: linear regression
  - To create a model such that the error between the predicted value of the target variable and the actual value is as small as possible
    - "fitting" the time series data (training data)

Values, TS patterns

- The primary difference (except **causal** models)
  - In time series models, early observations of the "target" variable are used to predict future values (no X variables, only Y)
  - Whereas other types of models use other variables as predictors

Forecasting Methods

There are two types of forecasting methods for TS data :

Qualitative : This type uses the opinion of experts to predict future event subjectively. Such methods are often used when the historical data is either not available or scarce

Example :

1. New product launch where no historical data is available. In such a case for forecasting the company has to rely on expert opinion
2. Another case where one needs to predict if and when new technologies will be discovered and adopted

The commonly used techniques are :

- Subjective curve fitting
- **Delphi method**
- Time independent technological comparisons

They are frequently called judgmental forecasting methods

Qualitative Technique: Delphi Methods

- The Delphi method is based on the assumption that group judgments are more valid than individual judgments.
- Predictions and reasoning from experts are collected and summarized by a moderator.
- Experts are now told the predictions of other experts and also provided feedback by the moderator
- Experts revise as they see fit
- After at least two rounds of feedback the forecasts are combined.
- The Delphi method was developed at the beginning of the Cold War to forecast the impact of technology on warfare – **relevant for policy making etc.**

Advantages of Qualitative Forecasts

- Takes advantage of the users' domain knowledge
- Can be used when we have no history, e.g., when a new product with no past analogue is introduced
- Can be used when available data is informal or qualitative and quantitative data is missing or sparse

Limitations of Qualitative Forecasts

- Impractical or too expensive when thousands of forecasts are required on a regular basis
  - Airlines passenger forecasts
  - Inventory planning
  - Economic Planning

- Expert judgement used for forecasting is subjective
  - It is lost when the expert leaves

- Expert judgement is often biased due to
  - Recency : recent occurrences are given greater weight
  - Risk aversion; human forecasters consistently under-forecast cancellation rates to avoid overbooking;
  - Hopes and fears of the expert
  - Confirmation bias

Famous Forecasts

The atom bomb will never go off - and I speak as an expert in explosives

Admiral William Leahy in 1945.

Heavier-than-air flying machines are impossible

Lord Kelvin, (1895) British mathematician and physicist

I think there is a world market for maybe five computers

Thomas Watson, chairman of IBM, 1943 on seeing the first mainframe computer

There is no reason for any individual to have a computer in their home

Ken Olson, 1977, Chairman and CEO DEC

"Stocks have reached what looks like a permanently high plateau.“

Irving Fisher, Professor of Economics, Yale University, 1929

The foolish idea of shooting for the moon is an example of the absurd length to which vicious specialization will carry scientists working in thought-tight compartments.

A.W. Bickerton (1926) Professor of Physics and Chemistry, Canterbury College, New Zealand

This 'telephone' has too many shortcomings to be seriously considered as a means of communication.

The device is inherently of no value to us." --Western Union internal memo, 1876

TS Forecasting Methods contd..

Quantitative : This type analyzes historical data in attempt to predict future values of a variable of interest.

Quantitative models are of two types a) Univariate Models & b) Causal Models

A **univariate (self projecting)** forecasting model predicts future values of a time series solely on the basis of the past values of time series. Hence this method is most useful when the conditions are expected to remain the same . It is not very useful in forecasting the impact of changes in management policies

Example : The univariate model can be used to predict future sales if the management continues to use the same marketing strategy. However the technique may not be useful to predict sales if there is an increase in price, increased advertising expenditure or if a new advertising campaign is launched

TS Forecasting Methods contd..

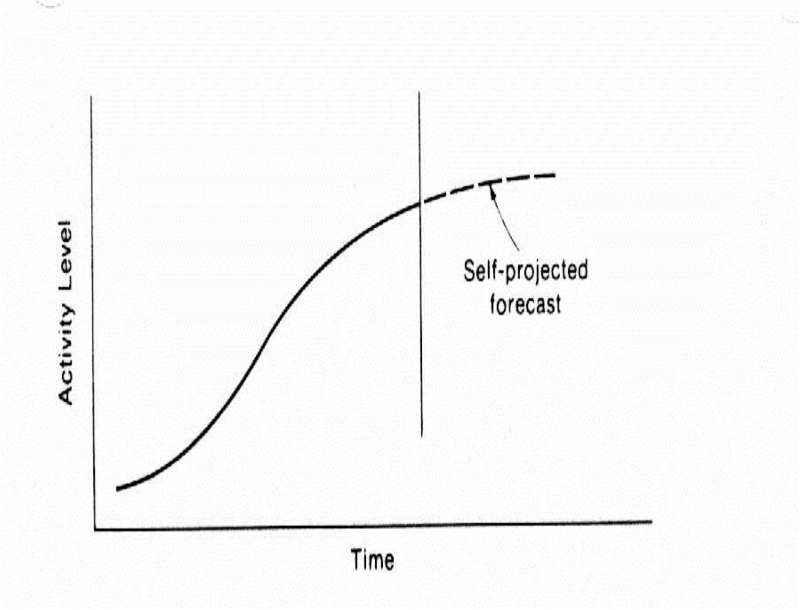
The **causal (cause & effect)** forecasting model involves identification of other variables (independent X variables) that are related to the variable to be predicted (dependent variable Y)

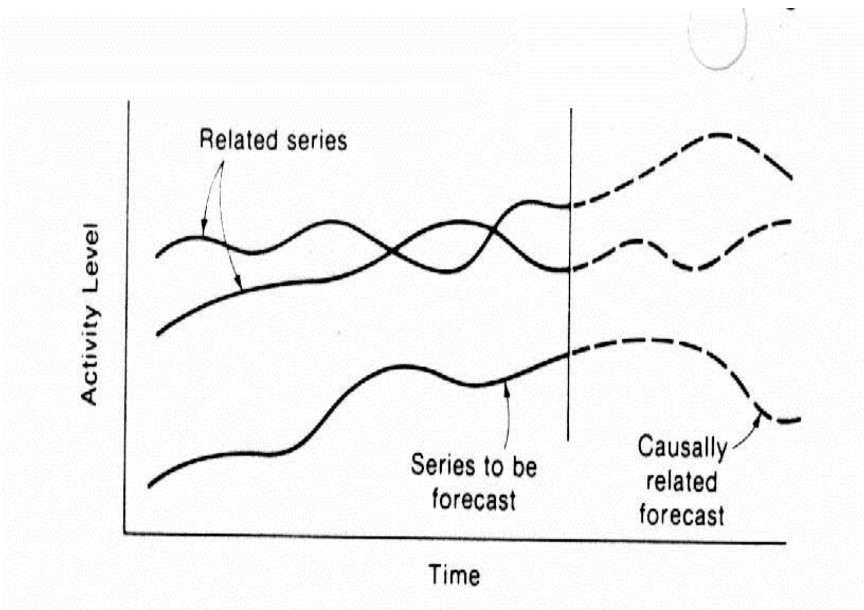
Example : The sales of a product might be related to the price of the product, advertising expenditure, price of the competitor and so on. The forecaster's job is to estimate statistically the functional relationship between sales and independent variables. Having determined this relationship, the forecaster would use the predicted future values of the independent variables (price of the product, advertising expenditure, price of the competitor etc.) to predict the future values of sales

In the business world, causal models are advantageous because they allow management to evaluate the impact of various alternative policies. For example management might wish to predict how various price structures and levels of advertising expenditure will affect sales.

Summary of Quantitative Methods

- Self-projecting approach (univariate)
- Cause-and-effect approach (multivariate)





Summary of Quantitative Methods (Cont.)

- Self-projecting approach
  - Advantages
    - Quickly and easily applied
    - A minimum of data is required
    - Reasonably short-to medium-term forecasts
    - They provide a basis by which forecasts developed through other models can be measured against
  - Disadvantages
    - Not useful for forecasting into the far future
    - Do **not** take into account external factors
- Cause-and-effect approach
  - Advantages
    - Bring more information
    - More accurate medium-to long-term forecasts
  - Disadvantages
    - Forecasts of the explanatory time series are required

Approaching Time Series Analysis

- There are many, many different time series techniques.
- It is usually impossible to know which technique will be best for a particular data set.
- It is customary to try out several different techniques and select the one that seems to work best.
- To be an effective time series modeler, you need to keep several time series techniques in your "tool box."

Limitations of Forecasting

- Many managers have an unrealistic attitude towards forecasting – some of are pessimistic about what can be accomplished while others are unduly optimistic
- The future is neither completely knowable nor totally obscure. High quality forecasts should be seen as having a very direct impact on "bottom line". Some uncertainty (which cannot be dealt with by forecasting methods) will inevitably remain
- Some quantities are notoriously more difficult to forecast that others
- Forecasting techniques are not capable of eliminating uncertainty about the future, but usually they do a good job in predicting potential "trends" and "characterizing" residual uncertainties

The Three Principles of Forecasting

- Forecasts are always wrong!
  - Forecasting is complex and forecasts are always subject to some error

Forecast Made → Event Realized

  - Very important to have a measure of the forecast uncertainty when it is made
  - Always necessary to monitor the error, i.e., difference between forecast and actual

The Second Principle of Forecasting

- Forecasts at a higher level of granularity are (almost always) more accurate than forecasts at a lower level of granularity
  - Sales forecasts for SKUs are likely to be worse than sales forecasts for brands; sales forecasts for brands are likely to be worse than sales forecasts for products
  - Sales forecasts for stores are likely to be worse than sales forecasts for cities; sales forecasts for cities are likely to be worse than sales forecasts for sales territories
  - It is important to determine the level(s) at which data should be aggregated so that the forecasts are good & meaningful

The third principle of forecasting

- Near term forecasts are (almost always) more accurate than distant term forecasts
  - 1 week ahead forecasts will usually be better than 1 month ahead forecasts;
  - 1 month ahead forecasts will usually be better than 2 month ahead forecasts;
  - 2 month ahead forecasts will usually be better than 1 year ahead forecasts
  - It is important to determine the forecast horizon(s) required for operational, tactical and strategic decisions

Evaluating Methods

- Forecasting method is selected - many times by intuition, previous experience, or computer resource availability

1. Divide the data into two parts - an initialization part and a test part
2. Use the forecast technique to determine the fitted values for the initialization data set
3. Use the forecast technique to forecast the test data set and determine the forecast errors
4. Evaluate errors (MAD, MPE, MSE, MAPE)
5. Use the technique, modify, or develop new model

Forecasting Accuracy

- We need a way to compare different time series techniques for a given data set.
- Four common techniques are the:
  - mean absolute deviation, $$\text{MAD} = \sum_{i=1}^{n} \frac{\left|\text{Y}_i - \hat{\text{Y}}_i\right|}{n}$$
  - mean absolute percent error, $$\text{MAPE} = \frac{100}{n} \sum_{i=1}^{n} \frac{\left|\text{Y}_i - \hat{\text{Y}}_i\right|}{\text{Y}_i}$$
  - the mean square error, $$\text{MSE} = \sum_{i=1}^{n} \frac{\left(\text{Y}_i - \hat{\text{Y}}_i\right)^2}{n}$$
  - root mean square error. $$\text{RMSE} = \sqrt{\text{MSE}}$$

Two types of forecasts are considered : a) Point Forecast b) Prediction Interval Forecast

What do these Measure?

- Bias – does the forecast tend to underestimate or overestimate the actual

- MAE – measures the error, i.e., the gap between forecasts and actuals

- MSE – also measures the error
  - Strongly penalizes large errors

- RMSE – scales down the MSE

- MAPE – measures the error as a percentage

Forecasting Philosophies

- Forecasting "Philosophies":
  - Some forecasting systems use many different models to obtain each forecast and then determines the best according to some statistical model
  - Others use a single model
  - Some combine forecasts from different models

- Classes of Models
  - Regression
  - Time series

TYPE OF MODELS

Quantitative Forecasting Techniques

Causal

- Regression Analysis

Univariate

- Time Series Regression (t is the independent variable)
- Classical Decomposition Methods
- Exponential Smoothing
- Box-Jenkins Methods

Trend Models (Time Series Regression)

$y_t = TR_t + \varepsilon_t$

Where y_t = value of the time series in period t

TR_t = the trend in time period t

ε_t = the error term in time period t

No Trend : $TR_t = \beta_0$

Linear Trend : $TR_t = \beta_0 + \beta_1 t$

Quadratic Trend : $TR_t = \beta_0 + \beta_1 t + \beta_2 t^2$

The pth – order polynomial trend model is :

$$TR_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \ldots\ldots\ldots\ldots\ldots + \beta_p t^p$$

Application

The State University Credit Union, a savings institution open to faculty and staff of the university handles savings accounts and makes loans to members.

In order to plan its investment strategies, the credit union requires both point predictions and prediction intervals of monthly loan requests (in '000 dollars) to be made by the faculty and staff in future months

Loan Requests ('000s)

Month	Year 1	Year2
Jan	297	808
Feb	249	809
Mar	340	867
Apr	406	855
May	464	965
Jun	481	921
Jul	549	956
Aug	553	990
Sep	556	1019
Oct	642	1021
Nov	670	1033
Dec	712	1127

Regression Statistics	
Multiple R	0.989028
R Square	0.978177
Adjusted R Square	0.977185
Standard Error	39.67949
Observations	24

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	261.1232	16.71897	15.61838	2.17E-13	226.4502	295.7962	226.4502	295.7962
X Variable 1	36.74348	1.170084	31.40242	9.08E-20	34.31687	39.17008	34.31687	39.17008

Use the data and do the necessary analysis to help the credit union in their decision making. Can you improve upon the above result?

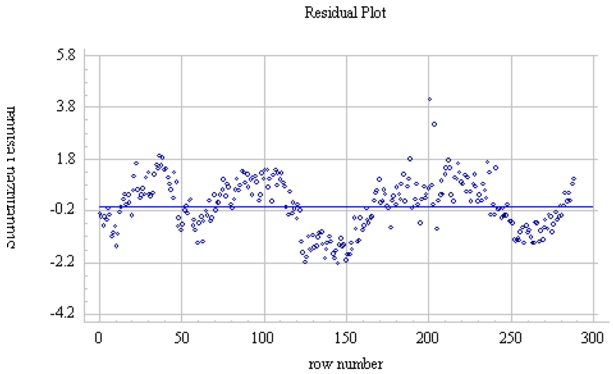
Detecting Autocorrelation

The validity of the regression method requires that the errors are independent. However when time series data are being analyzed this assumption is often violated. It is quite common that the time ordered error terms to be auto correlated

Just as correlation (r) measures the extent of linear relationship between two variables; **Auto-correlation** measures the linear relationship between lagged values of a time series. It can be of two types : positive and negative

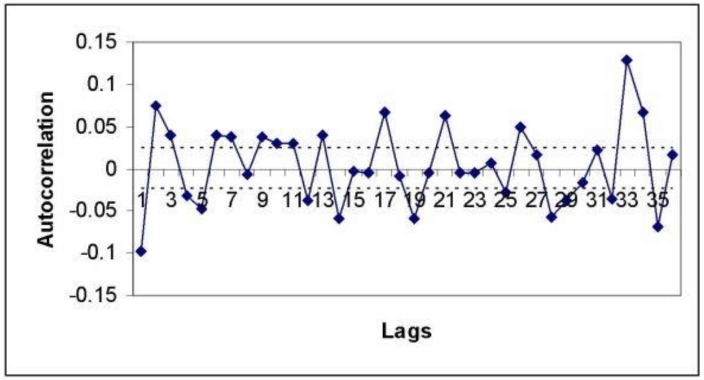
(exactly like simple correlation)

Positive Autocorrelation



Cyclical Pattern

Negative Autocorrelation



Alternating Pattern

Detecting Autocorrelation

Error terms occurring over time have **positive auto-correlation**

1) if a positive error term in time period t tends to produce, or be followed by another positive error term in time period t+k ,

& have **negative auto-correlation**

2) if a negative error term in time period t tends to produce, or be followed by another negative error term in time period t+k

+ Autocorrelation : Error sign + + + + - - - + + + - - - - - + + + +

- Autocorrelation : Error sign + - + - - + - + - + + - + - + - + +

Random : Error sign + - + + + - + - - + - + - - +

If the error term $\epsilon_t = \Phi_1 \epsilon_{t-1} + a_t$ can be expressed in this form where ϵ_t = error term in time period t , Φ_1 = correlation co-efficient and a_1 , a_2 , ... are i.i.d Normal with mean 0 and a variance independent of time then there exists **first-order autocorrelation**

Durbin-Watson Test for Autocorrelation Detection

DURBIN–WATSON TEST: (FIRST-ORDER) POSITIVE AUTOCORRELATION

The Durbin–Watson statistic is

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

where $e_1, e_2, \ldots, e_n$ are the time-ordered residuals.

Consider testing the null hypothesis

H_0: The error terms are not autocorrelated

versus the alternative hypothesis

H_a: The error terms are positively autocorrelated

Durbin and Watson have shown that there are points (denoted $d_{L,\alpha}$ and $d_{U,\alpha}$) such that if α is the probability of a Type I error, then

1. If $d < d_{L,\alpha}$, we reject H_0.
2. If $d > d_{U,\alpha}$, we do not reject H_0.
3. If $d_{L,\alpha} \le d \le d_{U,\alpha}$, the test is inconclusive.

DURBIN–WATSON TEST: (FIRST-ORDER) NEGATIVE AUTOCORRELATION

Consider testing the null hypothesis

H_0: The error terms are not autocorrelated

versus the alternative hypothesis

H_a: The error terms are negatively autocorrelated

Durbin and Watson have shown that based on setting the probability of a Type I error equal to α, the points $d_{L,\alpha}$ and $d_{U,\alpha}$ are such that

1. If $(4 - d) < d_{L,\alpha}$, we reject H_0.
2. If $(4 - d) > d_{U,\alpha}$, we do not reject H_0.
3. If $d_{L,\alpha} \le (4 - d) \le d_{U,\alpha}$, the test is inconclusive.

DURBIN–WATSON TEST: (FIRST-ORDER) POSITIVE OR NEGATIVE AUTOCORRELATION

Consider testing the null hypothesis

H_0: The error terms are not autocorrelated

versus the alternative hypothesis

H_a: The error terms are positively or negatively autocorrelated

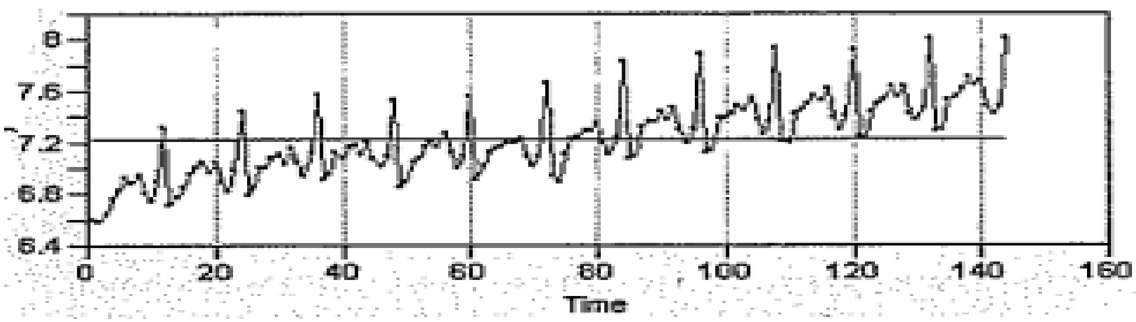
Durbin and Watson have shown that, based on setting the probability of a Type I error equal to α,

1. If $d < d_{L,\alpha/2}$ or if $(4 - d) < d_{L,\alpha/2}$, we reject H_0.
2. If $d > d_{U,\alpha/2}$ and if $(4 - d) > d_{U,\alpha/2}$, we do not reject H_0.
3. If $d_{L,\alpha/2} \le d \le d_{U,\alpha/2}$ or $d_{L,\alpha/2} \le (4 - d) \le d_{U,\alpha/2}$, the test is inconclusive.

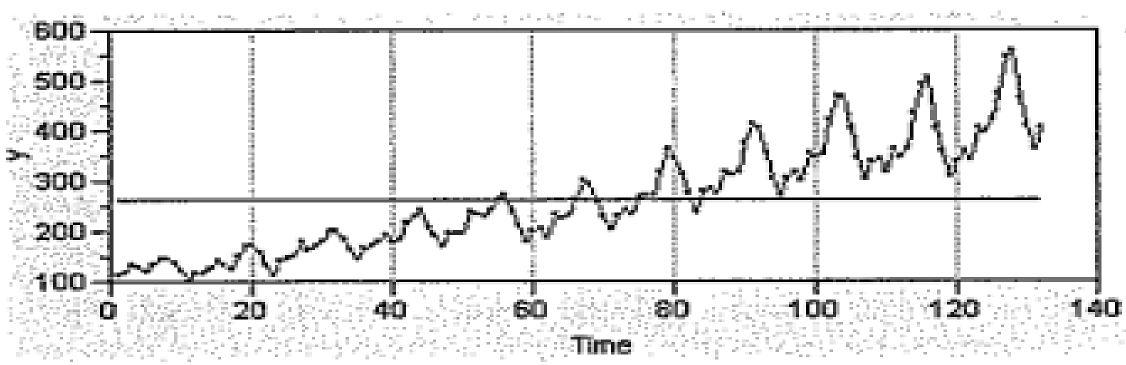
Time Series with Seasonal Variation

Two Types of seasonal variation for Time Series :

1) Constant Seasonal Variation : If the magnitude of seasonal swing does not depend on the level of the time series



2) Increasing Seasonal Variation : If the magnitude of seasonal swing depends on the level of the time series



In such case it is a common practice to convert **increasing** variation to **constant** variation using the Transformation

$y_t^* = y_t^\lambda$

where $0 < \lambda < 1$

As λ approaches 0 one uses $y_t^* = \ln y_t$

Modeling Seasonal Variation By Using Dummy Variables

While analyzing a time series that exhibits constant seasonal variation, a model of the following form is often used:

$$y_t = TR_t + SN_t + \varepsilon_t$$

where

y_t = the observed value of the time series in time period t
TR_t = the trend in time period t
SN_t = the seasonal factor in time period t
ε_t = the error term (irregular factor) in time period t

One way to model seasonal patterns is to employ **dummy variables.** Assuming there are L seasons (months, quarters etc.) per year the seasonal factor SN_t is expressed as

The seasonal factor expressed using *dummy variables* is

$$SN_t = \beta_{s1}x_{s1,t} + \beta_{s2}x_{s2,t} + \cdots + \beta_{s(L-1)}x_{s(L-1),t}$$

where $x_{s1,t}, x_{s2,t}, \ldots, x_{s(L-1),t}$ are dummy variables that are defined as follows:

$$x_{s1,t} = \begin{cases} 1 & \text{if time period } t \text{ is season 1} \\ 0 & \text{otherwise} \end{cases}$$

$$x_{s2,t} = \begin{cases} 1 & \text{if time period } t \text{ is season 2} \\ 0 & \text{otherwise} \end{cases}$$

$$\vdots$$

$$x_{s(L-1),t} = \begin{cases} 1 & \text{if time period } t \text{ is season } L-1 \\ 0 & \text{otherwise} \end{cases}$$

Modeling Seasonal Variation By Using Dummy Variables contd

When L = 12, we have monthly data. If the trend is linear the model becomes :

$$
\begin{aligned}
y_t &= TR_t + SN_t + \varepsilon_t \\
&= \beta_0 + \beta_1 t + \beta_2 M_1 + \beta_3 M_2 + \cdots + \beta_{12} M_{11} + \varepsilon_t
\end{aligned}
$$

Application

Traveler's Rest Inc operates 4 hotels in a Midwestern city. The analysts in the operations division of the organization were asked to develop a model that could be used to obtain short-term forecasts (up to one year) of the number of occupied rooms in the hotels. These forecasts were needed by various personnel to assist in decision making with regard to hiring additional help during the summer months, ordering materials that have long delivery lead times, budgeting of local advertising expenditure and so on.

Monthly average of hotel room occupancy y_t for 14 years have been provided. Explore the pattern of y_t and y_t^{λ} for various values of λ. Fit a seasonal model using dummy variables.

Smoothing Methods

Smoothing

- Respond to the most recent behavior of the series
- Employ the idea of weighted averages
- They range in the degree of sophistication

1. Simple Averages - quick, inexpensive (should only be used on stationary data)
2. Moving Averages - a constant number specified at the outset and a mean computed for the *most recent observations* - such as a 3 or 4 period moving average.
   - Works best with stationary data.
   - The larger the order of the moving average, the greater the smoothing effect. Larger *n* when there are wide, infrequent fluctuations in the data.
   - By smoothing recent actual values, removes randomness.

Smoothing Methods - Formula

- Simple average
  - m ~ most recently observation

$$\hat{Y}_{t+1} = \frac{Y_t + Y_{t-1} + .. + Y_{t-m+1}}{m}$$

- Moving average Smoothing (Trailing moving average)
  - w ~ window / width or interval

$$\hat{Y}_{t+1} = \frac{Y_t + Y_{t-1} + .. + Y_{t-w+1}}{w}$$

- Simple Exponential Smoothing
  - Smoothing parameter $0 < \alpha < 1$

$$\hat{Y}_{t+1} = \alpha Y_t + \alpha(1-\alpha) Y_{t-1} + \alpha(1-\alpha)^2 Y_{t-2} + \ldots +$$

Moving Averages

Observation	Demand	3 Period Moving Average
1	7	-
2	14	-
3	11	-
4	19	10.6667
5	9	14.6667
6	8	13
7	12	12
8	11	9.6667
9	7	10.3333
10	10	10
11	10	9.3333
12		9

- Simply forecast the next period observation to be the average of m previous observations
- Underlying assumption – the time series has a constant level (over the long term)
  - Very bad for time series with trend or seasonality
- How large should m be?
  - Large m (say 20) tends to smooth out the variability
  - Small m (say 5) useful for capturing recent changes

Exponential Smoothing

- Exponential Smoothing provides a forecasting method that is most effective when the components (trend and seasonal factors) of the time series may be changing over time
- More recent observations are weighted more heavily than more remote observations
- The unequal weighting is accomplished by using one or more smoothing constants, which determine how much weight is given to each observation

Exponential Smoothing

- If $y(1), y(2), \ldots,$ is a time series of observations in periods 1, 2, … and $F(2), F(3), \ldots$ the series of forecasts in periods 2, 3, … then

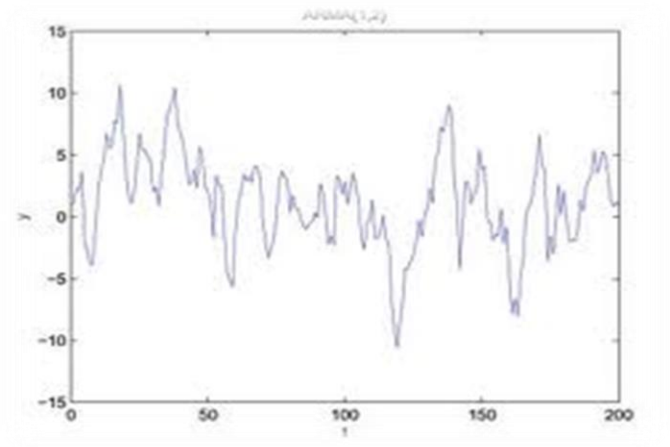
$$F(t+1) = \alpha y(t) + (1-\alpha)F(t)$$
$$= \alpha y(t) + (1-\alpha)\{\alpha\, y(t-1) + (1-\alpha)F(t-1)\}$$
$$= \alpha y(t) + \alpha(1-\alpha)y(t-1) + (1-\alpha)\,\{\alpha\, y(t-2) + (1-\alpha)F(t-2)\}$$
$$= \alpha y(t) + \alpha(1-\alpha)y(t-1) + \alpha(1-\alpha)\,\{y(t-2) + \ldots$$

- Underlying assumption – the time series has a constant or slowly changing mean
  - No trend or seasonality
- How large should α be?
  - If $\alpha = 1$, the forecast is the last observed value
  - If $\alpha = 0$, the forecast is the last forecast
  - Large α (say 0.8) tends to capture changes in the underlying process fast but are sensitive to noise (random fluctuations)
  - A compromise is required; values between 0.1 and 0.25 often used

Other Exponential Smoothing Methods

- Holt's Method
  - Generalizes exponential smoothing
  - Can be used for series which show trend
  - Two smoothing parameters

- Winter's Method
  - Can be used for series which show trend and seasonality
  - Three smoothing parameters

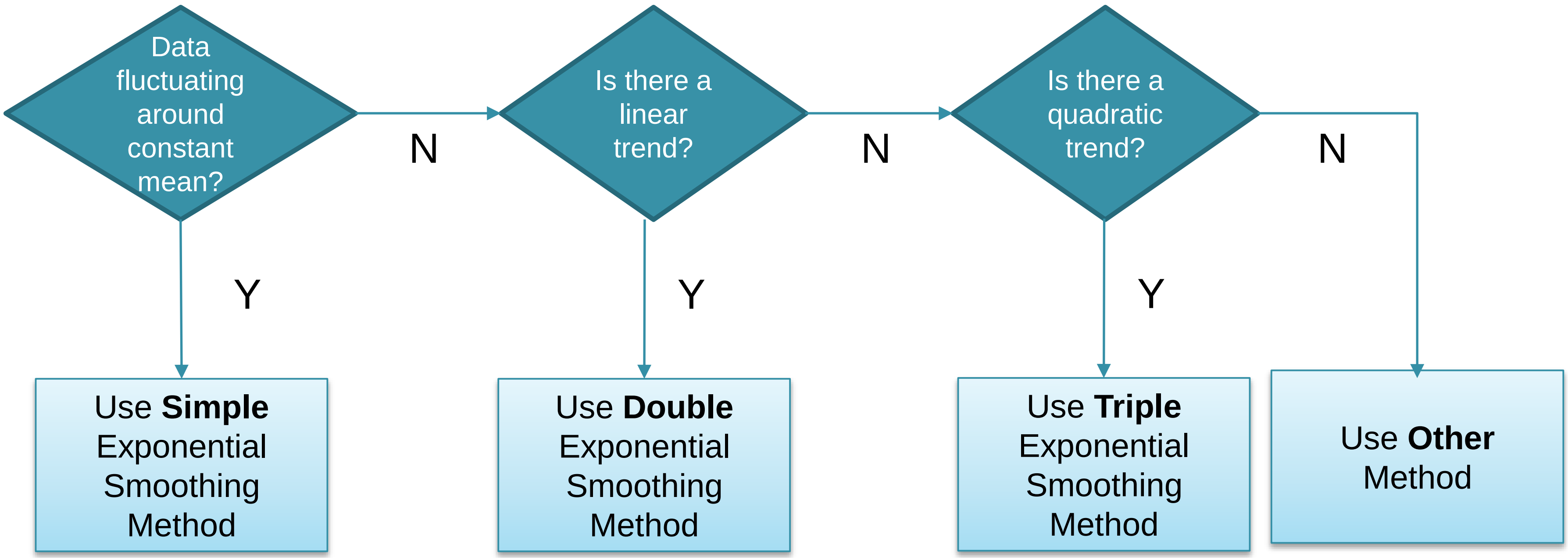


Exponential Smoothing

Observation	Demand	Exponential Smoothing (Damping Factor .3)
1	7	-
2	14	7
3	11	11.9
4	19	11.27
5	9	16.681
6	8	11.3043
7	12	8.99129
8	11	11.097387
9	7	11.0292161
10	10	8.20876483
11	10	9.462629449
12		

- All observations are **NOT** weighted equally
- Greater weight is given to more recent observations.
- Suppose that the damping factor is 0.3
- 70% weight is given to the most recent observation.
  - 70% * 30% = 21% is given to the 2nd most recent, etc.

Three types of Exponential smoothing methods



Simple Exponential Method

- Data fluctuating a constant mean, there is no trend .
- Consider **no trend equation**, where β_0 is changing over time.

$$Y_t = \beta_0 + \varepsilon_t$$

Step 1 : Find Initial estimate $a_0(0)$

$$a_{0(0)} = \bar{y} = \frac{\sum y_t}{n}$$

Step 2 : Compute $a_0(T)$ with an arbitrary α

$$a_{0(T)} = \alpha y_T + (1-\alpha) a_{0(T-1)}$$

Step 3 : Compute MSE/RMSE/MAD/MAPE

$$\text{MSE} = \sum_{i=1}^{n} \frac{\left(Y_i - \hat{Y}_i\right)^2}{n}$$

Step 4 : Select α that yield the lowest in step 3.

Double Exponential Smoothing

Double exponential smoothing can be used to apply unequal weightings when β_o and β_1 are slowing changing over time.

$$Y_t = \beta_0 + \beta_{1t} + \varepsilon_t$$

There are two versions of double exponential smoothing.

1) One parameter double exponential smoothing, using one smoothing constants. *(Brown's linear exponential smoothing)*

2) Holt two parameter exponential smoothing, using two smoothing constants.

$$\hat{y}_{t+h|t} = \ell_t + hb_t$$
$$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$$
$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$$

Classical Decomposition of Time Series

- **Trend** – does not necessarily imply a monotonically increasing or decreasing series but simply a lack of constant mean, though in practice, a linear or quadratic function is often to predict the trend;
- **Cycle** – refers to patterns or waves in the data that are repeated after approximately equal intervals with approximately equal intensity. For example, some economists believe that "business cycles" repeat themselves every 4 or 5 years;
- **Seasonal** – refers to a cycle of one year duration;
- **Random (irregular)** – refers to the (unpredictable) variation not covered by the above

Decomposition Method

Multiplicative Models

$$Y_t = TR_t \times SN_t \times CL_t \times IR_t$$

Additive Models

$$Y_t = TR_t + SN_t + CL_t + IR_t$$

Decomposition Method

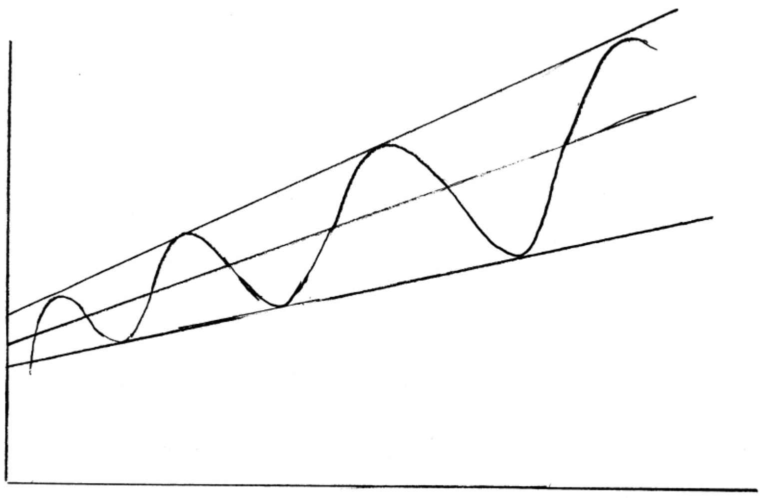
Find the estimates of these four components.

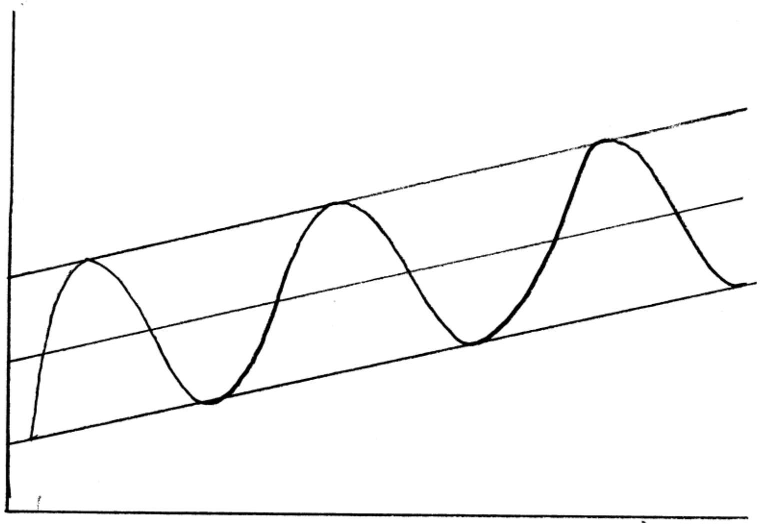
Multiplicative Models

$$Y_t = TR_t \times SN_t \times CL_t \times IR_t$$

Additive Models

$$Y_t = TR_t + SN_t + CL_t + IR_t$$





Multiplicative Decomposition

$$Y_t = TC_t \times SN_t \times IR_t$$

Examples:

Quarterly sales from 2010 to 2013

Period (t)	Year	Quarter	Sales
1	1	1	72
2		2	110
3		3	117
4		4	172
5	2	1	76
6		2	112
7		3	130
8		4	194
9	3	1	78
10		2	119
11		3	128
12		4	201
13	4	1	81
14		2	134
15		3	141
16		4	216

Step 1 : Estimation of seasonal component (SN_t)

- $Y_t = TC_t \times SN_t \times IR_t$

$$\therefore \qquad \hat{S}N_t = \frac{Y_t}{TC_t \times IR_t}$$

Calculating Center Moving Average

Period (t)	Year	Quarter	Sales	MA (4)	CMA
1	1	1	72		
2		2	110		
3		3	117	T=2.5 117.75	T=3 118.25
4		4	172	T=3.5 118.75	119
5	2	1	76	119.25	120.875
6		2	112	122.5	125.25
7		3	130	128	128.25
8		4	194	128.5	129.375
9	3	1	78	130.25	130
10		2	119	129.75	130.625
11		3	128	131.5	131.875
12		4	201	132.25	134.125
13	4	1	81	136	137.625
14		2	134	139.25	141.125
15		3	141	143	
16		4	216		

Calculating SNt

Sales	MA (4)	CMA	SN(t)	average[SN(t)]	Exp[SN(t)]
72				0.605768189	0.606312789
110				0.918243068	0.919068591
117	117.75	118.25	0.989429175	0.991230171	0.992121312
172	118.75	119	1.445378151	1.481165703	1.482497308
76	119.25	120.875	0.628748707		0.606312789
112	122.5	125.25	0.894211577		0.919068591
130	128	128.25	1.013645224		0.992121312
194	128.5	129.375	1.499516908		1.482497308
78	130.25	130	0.6		0.606312789
119	129.75	130.625	0.911004785		0.919068591
128	131.5	131.875	0.970616114		0.992121312
201	132.25	134.125	1.49860205		1.482497308
81	136	137.625	0.588555858		0.606312789
134	139.25	141.125	0.949512843		0.919068591
141	143				0.992121312
216					1.482497308

S1. SN(3) = 117/118.25=0.9894

S2. Average SN(t)
= average(3,7,11,…)

S3. Normalizing factor
= 4 /sum(average(SNT)

S4. Expected SN()
= S2 * S3

	3.996407131
normalizing factor	1.000899025

Step 2 : Estimation of Trend/Cycle

- Define deseasonalized (or seasonally adjusted) series as

$$D_t = Y_t / \hat{SN}_t$$

Compute D(t)

Sales	MA (4)	CMA	SN(t)	average[SN(t)]	Exp[SN(t)]	D(t)
72				0.605768189	0.606312789	118.7505876
110				0.918243068	0.919068591	119.686388
117	117.75	118.25	0.989429175	0.991230171	0.992121312	117.9291268
172	118.75	119	1.445378151	1.481165703	1.482497308	116.0204468
76	119.25	120.875	0.628748707		0.606312789	125.3478425
112	122.5	125.25	0.894211577		0.919068591	121.8625041
130	128	128.25	1.013645224		0.992121312	131.0323632
194	128.5	129.375	1.499516908		1.482497308	130.8602714
78	130.25	130	0.6		0.606312789	128.6464699
119	129.75	130.625	0.911004785		0.919068591	129.4789106
128	131.5	131.875	0.970616114		0.992121312	129.0164807
201	132.25	134.125	1.49860205		1.482497308	135.5820337
81	136	137.625	0.588555858		0.606312789	133.5944111
134	139.25	141.125	0.949512843		0.919068591	145.7997817
141	143				0.992121312	142.119717
216					1.482497308	145.700096

Estimation of Trend/Cycle

- TC_t may be estimated by regression using a linear trend:

$$D_t = \beta_0 + \beta_1 t + \varepsilon_t$$

$$t = 1,\ 2,\ 3\ldots$$

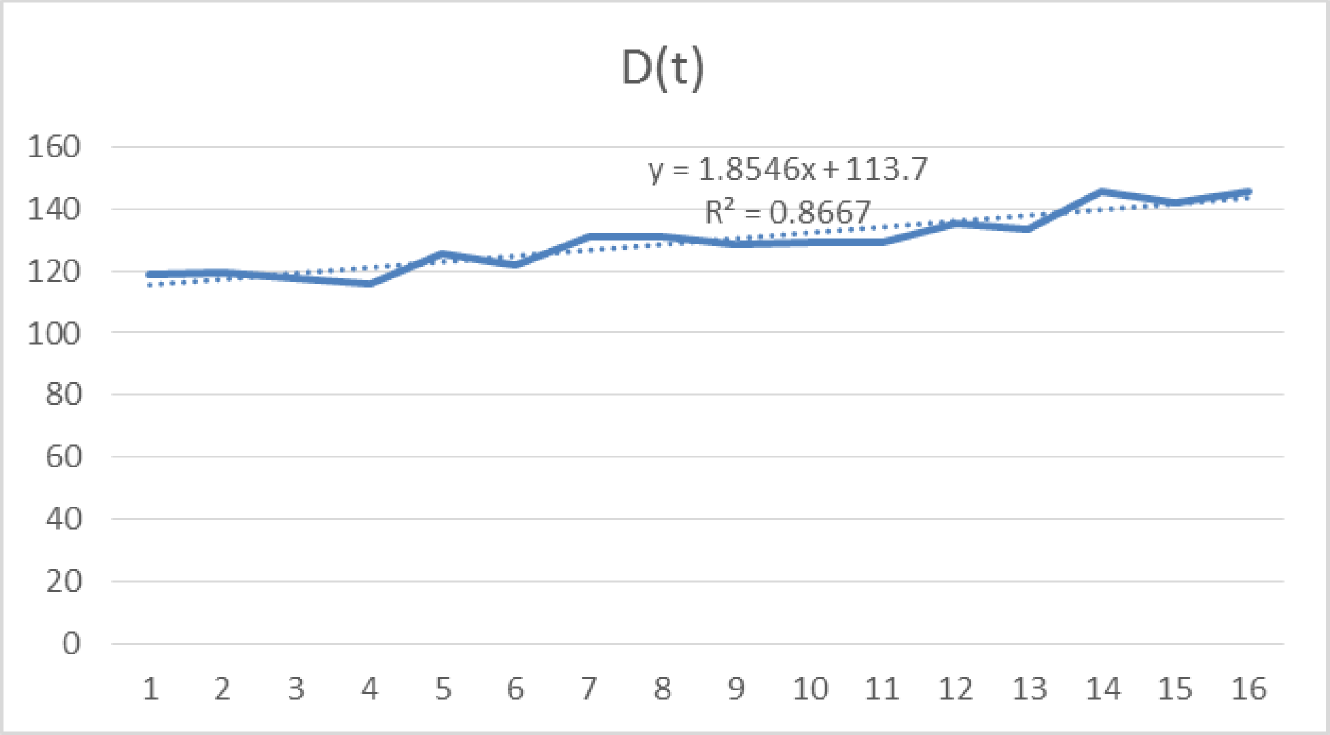
$$T\hat{C}_t = \hat{D}_t = b_0 + b_1 t \ ,$$

where b_0 and b_1 are least squares estimates of β_0 and β_1 respectively.

Calculating TC_t

Sales	MA (4)	CMA	SN(t)	average[SN(t)]	Exp[SN(t)]	D(t)
72				0.605768189	0.606312789	118.7505876
110				0.918243068	0.919068591	119.686388
117	117.75	118.25	0.989429175	0.991230171	0.992121312	117.9291268
172	118.75	119	1.445378151	1.481165703	1.482497308	116.0204468
76	119.25	120.875	0.628748707		0.606312789	125.3478425
112	122.5	125.25	0.894211577		0.919068591	121.8625041
130	128	128.25	1.013645224		0.992121312	131.0323632
194	128.5	129.375	1.499516908		1.482497308	130.8602714
78	130.25	130	0.6		0.606312789	128.6464699
119	129.75	130.625	0.911004785		0.919068591	129.4789106
128	131.5	131.875	0.970616114		0.992121312	129.0164807
201	132.25	134.125	1.49860205		1.482497308	135.5820337
81	136	137.625	0.588555858		0.606312789	133.5944111
134	139.25	141.125	0.949512843		0.919068591	145.7997817
141	143				0.992121312	142.119717
216					1.482497308	145.700096

Calculating TC_{t}.... cont'd



Step 3 : Computation of fitted values and out-of-sample forecasts

$$\hat{Y}_t = \hat{T}C_t \times \hat{S}N_t$$

Fitted Values

Period (t)	Year	Quarter	Sales	E(Y)= exp(TC) * exp(SN)
1	1	1	72	70.06
2		2	110	107.91
3		3	117	118.32
4		4	172	179.56
5	2	1	76	74.56
6		2	112	114.73
7		3	130	125.68
8		4	194	190.56
9	3	1	78	79.06
10		2	119	121.54
11		3	128	133.04
12		4	201	201.55
13	4	1	81	83.56
14		2	134	128.36
15		3	141	140.40
16		4	216	212.55

Measuring Forecast Accuracy :

Let $e_t = Y_t - \hat{Y}_t$ be the errors of forecast.

1) Mean Squared Error

$$MSE = \sum_{t=1}^{n} e_t^2 / n$$

$$RMSE = \sqrt{MSE}$$

2) Mean Absolute Deviation

$$MAD = \sum_{t=1}^{n} |e_t| / n$$

$$RMAD = \sqrt{MAD}$$

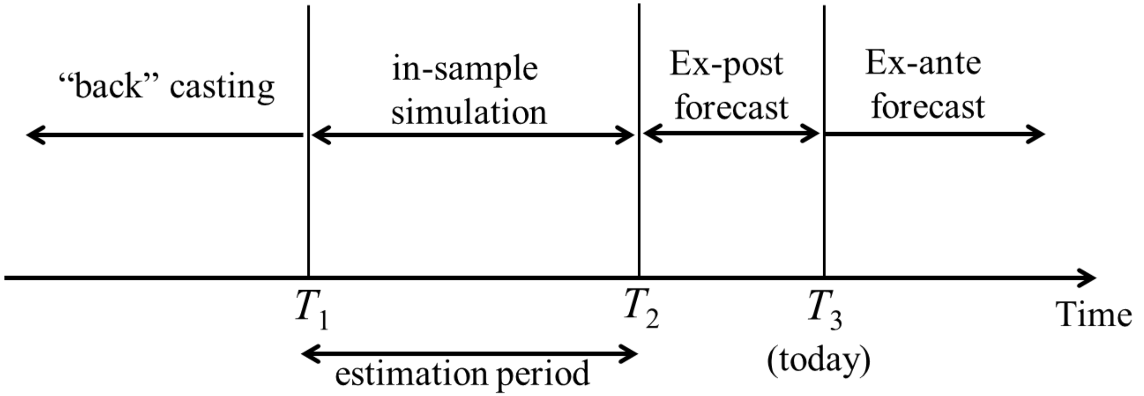
Model Checking

MSE	RMSE	MAD	RMAD`
190.91	13.82	46.28	6.80

Out-of-Sample Forecasts

1) Ex post forecast
   - Prediction for the period in which actual observations are available

2) Ex ante forecast
   - Prediction for the period in which actual observations are not available.



Summary

- Concept of time series data and its components.
- Concept of autocorrelations
- Uncovering patterns from time series data
- Key techniques
  - Time Series Regression
  - Smoothing methods
  - Decomposition methods

Data Issues

- How much data is required for forecasting?
  - Enough data to identify seasonal patterns (normally this should be at least 2 years)
  - For regression methods there should be at least 4 times as many data points as there are independent variables
  - Too much data may also be bad as old data tends to become “stale” or “irrelevant”
- Bad data should be marked and excluded

User Intervention

- It is necessary for users to “influence” forecasts to correct for
  - Changes in market conditions
  - Competitor actions
  - Political problems
  - Special events (conventions, strikes, etc.)
- Many forecasting systems allow users to add a number to a forecast or multiply it by a percentage
- Users may use existing products to “sponsor” new ones.
- For best forecasts statistical techniques should be combined with domain knowledge

Accuracy Measures

Time	Actual	Monthly Forecasts	Bias	Absolute	Squared	AbsolutePercentageE	
38749	977.5	991.882	-14.3816458	14.382	206.83	0.0147127	
38777	963.5	971.426	-7.925543963	7.9255	62.814	0.0082258	
38808	1032.5	974.965	57.53476989	57.535	3310.2	0.0557237	
38838	1098.75	1042.8	55.94696936	55.947	3130.1	0.0509187	
38869	1244	1092.49	151.5117438	151.51	22956	0.121794	
38899	1345	1240.23	104.7717919	104.77	10977	0.0778972	
38930	1493.13	1320.87	172.2566844	172.26	29672	0.1153666	
38961	1383.5	1486.74	-103.2430715	103.24	10659	0.0746246	
38991	1098.75	1360.65	-261.895459	261.9	68589	0.2383576	
39022	1093.13	1133.67	-40.54655036	40.547	1644	0.0370923	
39052	1157.5	1172.71	-15.21146997	15.211	231.39	0.0131417	
39083	1146.25	1177.87	-31.61601431	31.616	999.57	0.0275821	
Average			5.600183701	84.737	12703	0.0696198	
					112.71		

REFERENCE BOOK

- Forecasting, Time Series, and Regression by Bruce L. Bowerman, | Richard T O'Connell & Anne B Koehler, 4th Edition

APPENDIX - ARIMA

- Auto Regressive Integrated Moving Average
  - A sophisticated and very general class of methods
  - Includes exponential smoothing and many other methods as special cases
  - Tries to find complex patterns in the data
  - Requires a long data series – at least 50-60 data points
  - Requires skill and statistical training to use well
  - Very effective in the right hands

END