March 2003

Data Analysis and Statistical Model selection

SAP America, Inc.
3999 West Chester Pike
Newtown Square
PA 19073

Copyright

TABLE OF CONTENTS

Target Audience

This document is intended both for the novice and experienced Demand Planning Consultant/Professional. The reader is expected to have a basic understanding of Statistics.
This is not a research publication. It is a collection/summary of relevant information gathered from various sources, to meet the needs of the Demand Planning Professional. It will serve as an overview of the process and as a quick reference guide.
The contents of this paper are credited to a number of articles and information gathered from various sources in print and from the Internet. The list of the resources is listed in the Appendix. We acknowledge the work of these authors.

This is our attempt to bridge the gap between the Statistician and the Functional consultant.

Compiled By:
SCM Centre of Excellence - New Town Square

Doris Kimaru – SAP South Africa
Purvi Todai – SAP America
Velluva “Mo” Mohandas – SAP America

1. Introduction to Statistics

1.1. General

The well-known statement attributed to Mark Twain, there are 3 kinds of Lies – Lies, Damned Lies and Statistics - is often quoted by cynics who do not have faith in statistics.
Statistics is a set of methods that are used to collect, analyse, present, and interpret data. Studying a problem through the use of statistical data analysis usually involves four basic steps.

- Defining the problem
- Collecting the data
- Analysing the data
- Reporting the results

As Demand Planners who use Statistics as the basis of our forecast it is extremely important that we understand what statistics can and cannot do. The basis of our forecast and the assumptions on which the model is build has to be clearly understood.

In this discussion we are focusing on – Analyzing the data – stage only. In this stage, we do the data analysis and the model selection for the purpose of generating a forecast.

It is important to note the difference between forecasting and planning.

Forecasting: what one thinks ***will*** happen

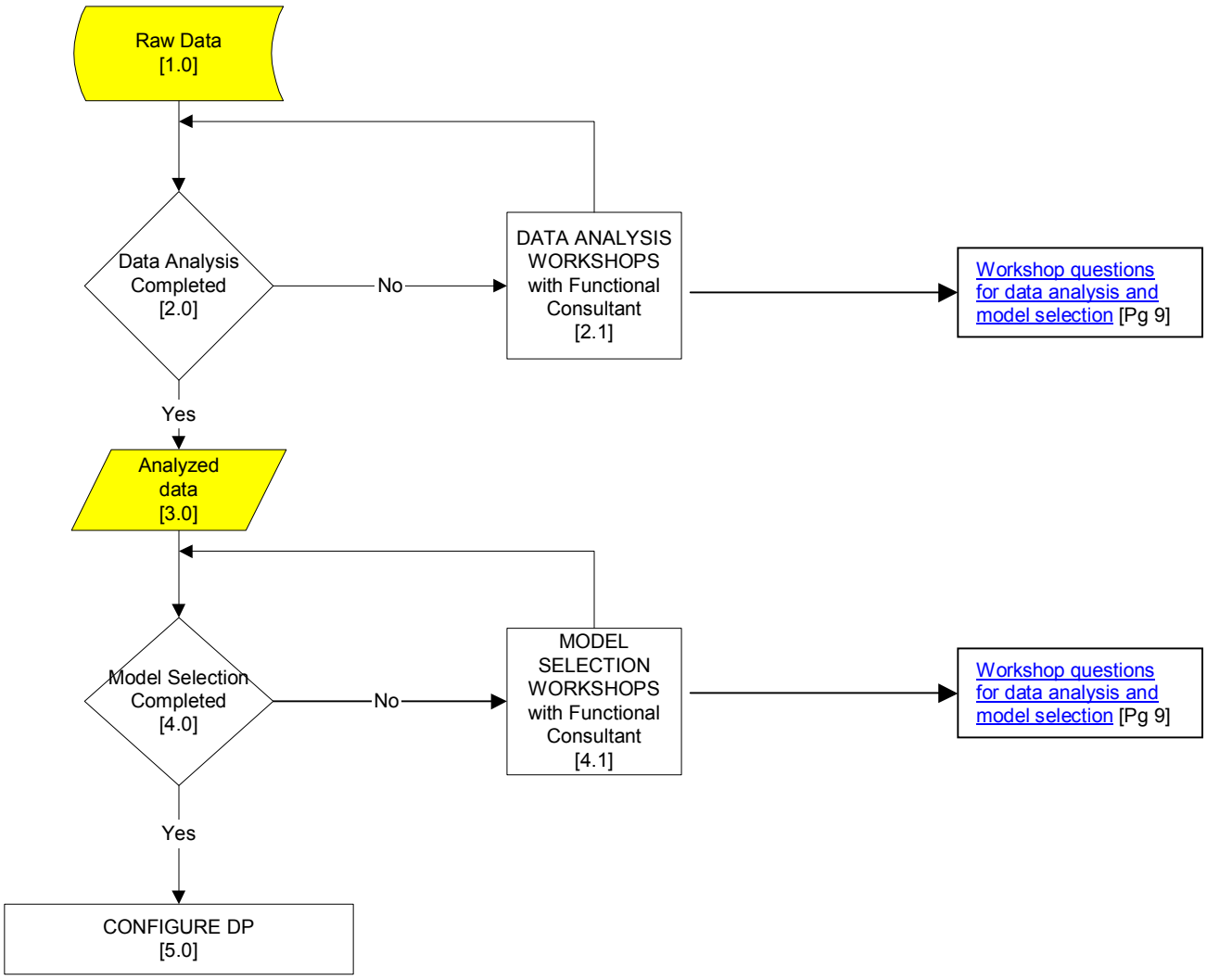
Planning: what one thinks ***should*** happen

"A forecast is an input into a plan"

The role of a forecaster can be summarized as:

Analyse the data to identify a pattern that can be used to describe the data – **DATA ANALYSIS**
Selecting the appropriate model that will allow extrapolation of the data pattern into the future – **MODEL SELECTION**

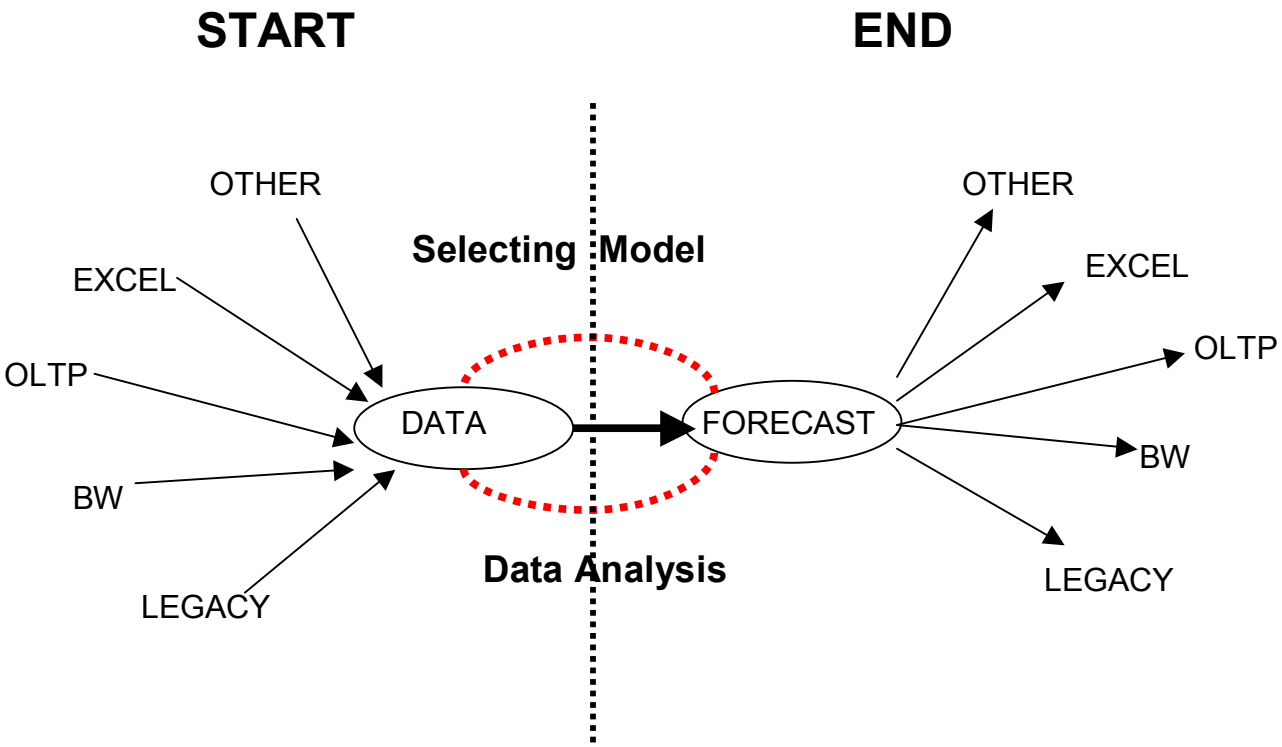
1.2. Overview of Forecasting Process



2. Data Analysis

2.1 Source of Raw Data

In APO, the data comes in raw format from the following sources, Legacy, OLTP, Excel, BW, and others.



2.2 Data Format

This data is usually in the form of daily, weekly or monthly buckets. Typically we analyse data at the lowest level, depending on the nature of the business. We then decide the level of aggregation to be used for forecasting ex: weekly vs Monthly. (Typically we don't recommend data to be run at the daily level for long term forecasting, for reasons of sizing and performance. Also the forecast is typically not used at the daily level)

2.3 How much data should you use to prepare forecasts?

The amount of data used depends on:

- How quickly the data pattern changes
- Type of industry
- Type of product
- Model selected
- Forecast horizon
- Ex-post forecast results

2.4 Goals of Data Analysis

The primary goal of data analysis is to maximize the analyst's insight into a data set and into the underlying structure of a data set, while providing all of the specific items that an analyst would want to extract from a data set, such as:

- A good-fitting, prudent model that suits the data.
- A list of outliers
- A sense of robustness about conclusions
- Estimates for parameters
- Uncertainties for those estimates
- A ranked list of important factors
- Conclusions as to whether individual factors are statistically significant

2.5 Workshop Overview

The purpose of this workshop is to determine the data that the customer would like to use for forecasting process. It is a process to determine data requirements & availability by selecting products to be used in the Demand Planning process.

2.5.1 Attendees

Client Strategic head of planning
Client Demand Planner
Client Head of IT
Client IT resource
SAP DP Team Lead
SAP DP Consultant

2.5.2 Workshop questions for data analysis and model selection

Data is analysed using statistics. Workshops will need to be held with the customer to establish the following:

(These questions are largely from the SAP APO online help documentation)

<1> How many products do you wish to forecast? All products? Which products? What are the most critical products from the forecasting perspective? [Pg 33-34]

<2> What is the granularity of the data (dd/ww/mm)? Is the historical data available in days, weeks or months? [Pg 34]

<3> What is the Frequency of forecasting. How often do you plan? – Weekly? Monthly? Quarterly [Pg 34]

<4> How much Data is required? How much data is to be used to prepare for the forecast? [Pg 34]

<5> Is the present market situation similar enough to the past situation that I can use existing historical data, or has there been some major structural or economic change that renders my historical data unsuitable for forecasting future demand? [Pg 35]

<6> If I have no historical data for a product, is it possible to forecast that product using the historical data of another product? [Pg 35]

<7> For which products do I already know the pattern of historical data? Or are the patterns of historical data clearly visible? [Pg 35]

2.6 Analysing the Data

Expert Selection begins by analysing the data. Once that analysis is completed, it generates an audit trail -- a report that shows the various mathematical methods the program has applied to the data, what the level of error is likely to be, the final choice of statistical modelling method, and the forecasted results.

The data is typically stored in:

- Microsoft Excel
- Microsoft Access
- Other software

Most of the available software provide you with complete flexibility to analyse and summarize data. Regardless of where your data originates, you can use all of the powerful features of any applications to create the summaries, reports, and charts you need.

What method should you use to analyse your sales data?

2.6.1 Plotting Data

Plotting data is necessary because it makes it easier to see patterns and trends in data that numbers alone can't show.
These patterns and trends further help to identify the appropriate model.

You can plot data using Microsoft excel and/or Microsoft access and/or other applications. [Pg 43]

2.6.2 Summarizing Data

Chances are, you have all the detailed data you need to make decisions, but it isn't always presented in a way that makes it easy to draw conclusions from it. For example, suppose what you want is the big picture: How is each product selling? Who is selling the most of each product?

From the same data, you can create several instant summaries, called **PivotTables**, to answer your questions. If you work with sales figures or other similar business data, Microsoft Excel can rapidly produce the summaries you want from the details you have.

If you're working on a sales report, you can create an overview of sales results to see how well sales representatives are doing and which products are selling the most for example. To do so, store

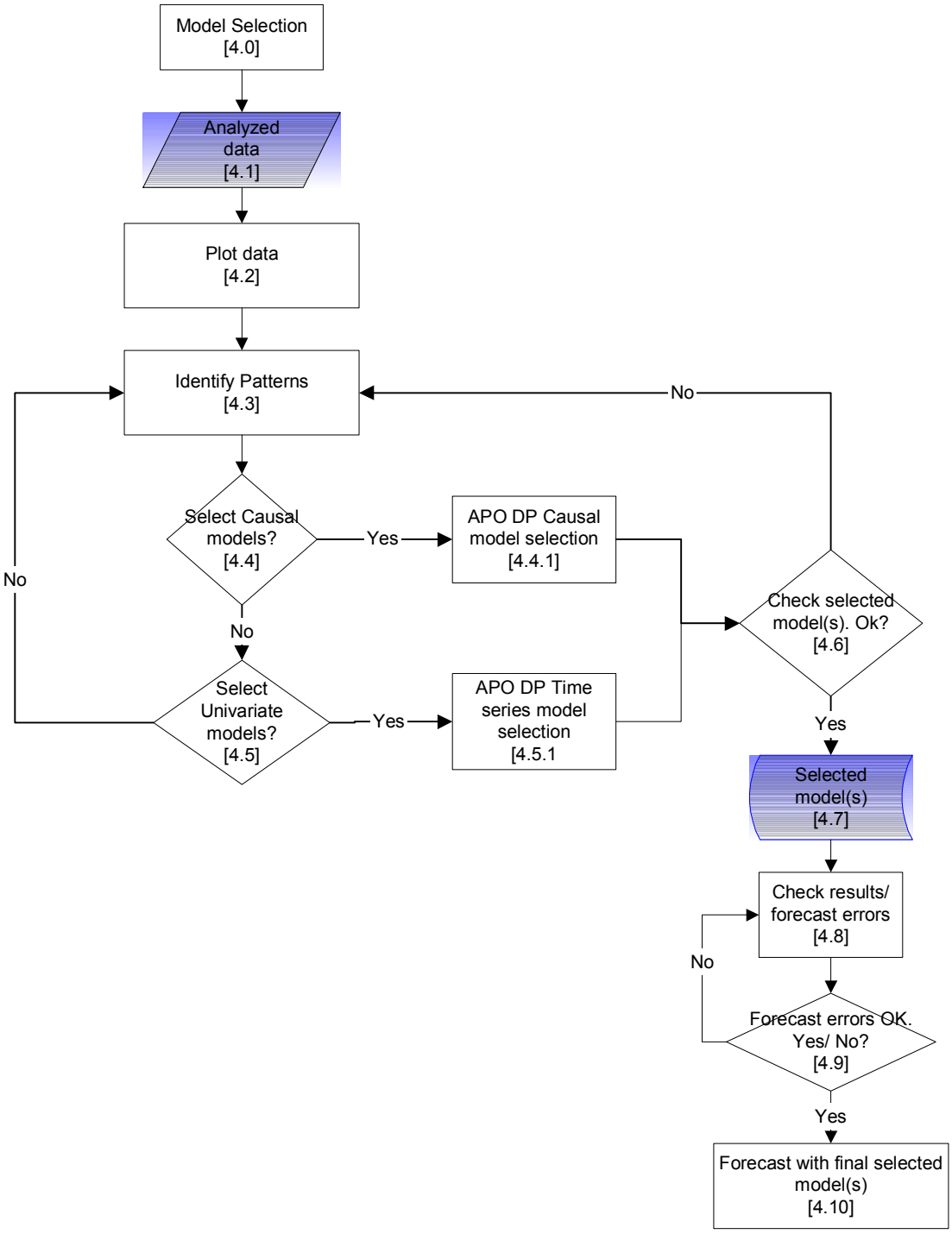
details about each order in a Microsoft Access sales database, and then switch to Microsoft Excel to summarize the data. Microsoft Excel can automatically create an interactive table, called a **PivotTable** that summarizes large amounts of data. The PivotTable stays current because whenever you change information in the sales database, it's automatically updated in the PivotTable as well.

3. Statistical Model Selection

There is no systematic approach for the identification and selection of an appropriate model, and therefore, the identification process is by trial and error.

Now that the data set has been analysed, the demand planner is faced with the task of selecting the RIGHT model that will generate the most accurate forecast.

3.1 Process Flow



3.2 Factors Influencing Model Selection

In general, selection of an appropriate technique or model selection can be guided by considering the following key factors about the forecasting situation.

Factors
Forecast Horizon [Pg 14]
Data Availability / Requirements [Pg 15]
Pattern of Past Data [Pg 15]
Level of Detail in the Forecast [Pg 15]
Cost Factors [Pg 15]
Accuracy [Pg 16]
Ease of Application [Pg 16]

Forecast Horizon:

Basically, one has to make sure that the technique allows you to pick up changes that might occur during the forecast time interval. For example,

- **Short-term: < 3 months**

In the short-term, seasonal fluctuations and randomness have little influence on sales volumes and purchasing decisions.

Short-term forecasts are mostly used in Operations Planning, determining inventory levels, work force levels and scheduling.

- **Medium-term: 3 months to 2 years**

Medium-term forecasts require that fluctuations of a medium-term nature (e.g., economic and competitive conditions) be accounted for by the technique. Since cyclical change and trend are important factors in this time frame, techniques such as regression analysis and time-series methods are useful.

Medium term forecasts would be used in budget planning & cost control, new product planning, capacity, warehouse & distribution planning and transportation contracts

- **Long-term: > 2 years**

With long-term horizons, the major consideration is with expected trends, as well as economic, competitive, and technological conditions, which can only be estimated subjectively. In many situations judgmental methods are employed.

Long-term forecasts would be used in capacity expansion, facility location, new market entry and workforce management.

Data Availability / Requirements:

Techniques differ by virtue of how much data is available and required to successfully employ the technique.

Pattern of past data:

The pattern of a product's previous sales history is an important factor to consider. While a commonly seen pattern is the trend, there are also cyclic and seasonal patterns to consider. Certain techniques are best suited for capturing the different patterns, such as Horizontal, Seasonal, Cyclical, Trend etc in the data.
The underlying data pattern typically encompasses one of the following. See the graphs in the Appendix 6.4 [Pg 38]

Horizontal Pattern (Constant)
Seasonal Pattern
Cyclical Pattern
Trend Pattern
Random or Irregular Pattern
A combination of the above

Level of detail:

Aggregate or Detailed Level Planning for Characteristics and Time.

- **Aggregate Level:**

If the planning is done at the aggregate level, the resulting forecast can be disaggregated down to the detail level based on proportional factors. At the higher levels the individual product patterns are not clearly visible.

- **Detail Level:**

On the other hand if the planning is done at the detail level, the patterns are visible. Here it is summed up to the next level.
When the forecast is disaggregated to the detail level, these patterns become more apparent.

Cost Factors

The selection of the model is effected by the following cost related factors

The time taken to develop a new model
The cost involved in additional data storage equipment
The time taken to do the modelling (Opportunity Cost)
The time taken to do a model selection

Accuracy

Different products demand different levels of accuracy. For example if we are forecasting aircrafts the accuracy levels are more critical than for example diapers.

Ease of Application

Sometimes only those models that are easily understood are used, though more complex models may be more accurate, for example econometric models.

3.3 Categories of Forecasting Methods

In many situations, there is rarely one best technique for selecting a model. It is not unusual to combine multiple forecasting techniques. The 3 major categories of forecasting methods are Quantitative, Qualitative and Unpredictable.

Elements of Forecasting methods	Quantitative	Qualitative	Unpredictable
Characteristics	Information about the past is available Can be quantified in the form of numerical data	Rely on Subjective assessments of people, using intuition based on experience	Little or no information is available
Assumption	Past Pattern will continue into the future	Experienced and knowledgeable person is available	
Techniques	Time Series & Causal	Delphi, Jury of executive	Crystal ball!!
Example	Predicting Monthly Sales	Predicting speed of Telecommunications in 2020	Predicting the discovery of a new, very cheap form of energy that produces no pollution.

3.4 Quantitative Methods

3.4.1 Time Series Models

"History repeats itself" is the central theme of quantitative time series forecasting A Time Series is a series of observations over time of some quantity of interest (a random variable).

The time series models, also know as Univariate forecasting techniques in APO Demand Planning consist of multiple strategies, which are explained, in the following table.

The various statistical models used in APO DP and their specific details are listed in the table below.

Time Series Models
Constant Models [Pg 17]
Trend Models [Pg 19]
Seasonal Models [Pg 21]
Seasonal Trend Models [Pg 21]
Auto Selections Models [Pg 22]
Croston Model [Pg 24]
Other Models [Pg 25]

APO DP has two choices of Auto selection of Statistical Models. However it is to be noted that forecasting can never be relegated completely to the automatic forecasting capability of the forecasting software. Not all-forecasting software is capable of recognizing and adjusting for "special events". For example, a one time special event occurring in October may well lead to a model that erroneously projects upward spikes in October months in the future.

3.4.1.1 Constant Models

Forecast Method		Forecast Strategy		Details	Observations
10 - Constant Models	10	Constant Models	A	Used for data that has no Trend or Seasonality.	Same as Strategy 11.
			B	Used when Demand varies slightly from the stable mean value.	
	11	Constant Model with First Order Exponential Smoothing	A	Used for data that has no Trend or Seasonality.	Same as Strategy 10.
			B	More recent data has more weight.	
			C	The user chooses the smoothing factor for the Basic (Alpha) value	Fairly good accuracy for Short Term (<= 3 months). Not

					recommended for Medium Term & Long Term.
			D	The forecast calculated using the historical value, the preceding forecast value & the Alpha smoothing factor.	Hard to find correct Alpha value.
			E	If data has Trends, the forecast will lag. This can be overcome by using 2nd order exponential smoothing. (Strategy 21)	Most widely accepted Time series.
					Strategy 11 (First Order Exponential Smoothing) does not apply to data have any underlying Trend or seasonality.
	12	Constant model with Automatic Alpha Adaptation (1st order)	A	Used for data that has no Trend or Seasonality.	
			B	The Alpha factor is adapted in every ex-post forecast.	
			C	The Alpha factor is based on Mean Absolute Deviation (MAD) & Error Total (ET) ranging from 0.05 to 0.90. The default Alpha factor is 0.3.	
			C	The Alpha factor is based on Mean Absolute Deviation (MAD) & Error Total (ET) ranging from 0.05 to 0.90. The default Alpha factor is 0.3.	
	13	Moving Average	A	Used for data that has no Trend or Seasonality.	Special emphasis word "equal" & "every".
			B	All Historical data gets **equal** weight	
			C	Calculates average of the	

				historical data in the time horizon.	
			D	The Average value is the forecast result for **every** period in the forecast horizon.	
	14	Weighted Moving Average	A	Every Historical value is weighted with a factor (From the Weighting Group)	Sum of weighting factors does not have to be 100%.
			B	No ex-post forecast calculated.	Accuracy of this model largely is dependant on user's choice of weighting factors.
			C	This model gives more weight to recent data and can therefore react more quickly to changes.	

3.4.1.2 Trend Models

Forecast Method		**Forecast Strategy**		**Details**	**Observations**
20 - Trend Models	20	Forecast with Trend Models	A	Used for data that has no Seasonality nor Constant	Same as Strategy 21
	21	Holt's method	A	Better accuracy for forecasting periods of less than 3 months.	Also known as Holt's Linear Trend method or Double exponential Smoothing.
			B	Smoothing factor Alpha, between 0 and 1.0 is assigned to the model.	Strategy 11 (First Order Exponential Smoothing) was extended by Holt's to allow forecasting of data with Trends.
			C	More weight for recent sales history is allowed, by increasing the value of	Uses 2 smoothing constants Alpha

				Alpha (say 0.8)	& Beta.
			D	Smaller Alpha (0.1) decreases the impact of recent data.	Does not work with Seasonal Patterns.
					Same as Strategy 20
	22	2nd Order exponential Smoothing	A	It is based on Linear Trend.	The key difference between the 1st order and 2nd order exp smoothing is based on data patterns. If the data series has a horizontal pattern,(no trend) 1st order is used. If it has trend, 2nd order exponential is used.
			B	Consists of 2 equations.	
			C	If, over several time periods, time series shows a change in the average value such that a trend pattern is revealed, first-order exponential smoothing produces forecast values that lag behind the actual values by one or several periods. You can achieve a more efficient adjustment of the forecast to the actual values pattern by using second-order exponential smoothing.	
	23	Trend model with Automatic Alpha Adaptation (2nd order)		Same as 22, except the system selects the value of Alpha.	Same automatic Alpha adaptation is used in Strategy 12 and 23
				The Alpha factor is generated by the system,	

				based on the MAD and ET	
				The Alpha values are between 0.05 and 0.90	

3.4.1.3 Seasonal Models

Forecast Method		Forecast Strategy		Details	Observations
30 - Seasonal Models	30	Forecast with Seasonal Models.		Uses 1st order Exponential smoothing	Same formula for 20,21,30,31, 40 & 41
				The Seasonal factor is Gamma. Here Alpha and Gamma factors are given values.(Both trend and seasonal variations are taken into account)	Holt's Method was extended by Winters to capture seasonality.
	31	Seasonal Trend based on the Winters method			Same as 30
	35	Seasonal + Linear Regression	A	Here Seasonal Indices are calculated and the seasonal data is removed, before linear regression is done on the remaining data.	
			B	The seasonal indices are added back after the LR is done.	
			C	Do NOT use if there are strong TREND patterns	
				Smoothing factor (PERMSO) is to be 1. If 0 is used, no seasonal influence is calculated	

3.4.1.4 Seasonal Trend Models

Forecast Method		Forecast Strategy		Details	Observations
40 - Seasonal Trend	40	Forecast with seasonal trend models		Here Alpha, Beta and Gamma values are used	Same formula for 20,21,30,31, 40 & 41

Models					
					Here the Alpha, Beta and Gamma factors are all given values
	41	First - Order Exponential Smoothing			Same as 40

3.4.1.5 Auto Selection Models

Forecast Method		**Forecast Strategy**		**Details**	**Observations**
50 - Auto Selection 1	50	Auto Selection 1	A	Used when there is no knowledge of patterns in historical data	Quicker than Auto Model Selection 2 (Strategy 56).
			B	Tests for Constant, Trend, Seasonal & Seasonal Trend.	Not as precise as Auto Model Selection 2 (Strategy 56).
			C	If no pattern detected, system uses Constant model	Shorter time to run in comparison with Auto Model Selection 2 (Strategy 56).
					Auto Selection Model 1 is used when data pattern is not obvious.
	51		A	Used if you **think** historical data has Trend pattern & NO other pattern	Special emphasis word "think".
			B	Regression Analysis is run on Historical data to check for trend	Faster than Auto Model selection 1 & 2, as limited tests carried out.
			C	If no Trend pattern detected, system uses Constant model	Vice-versa of strategy 52; Checks for Trend pattern.

	52		A	Used if you **think** historical data has Seasonal pattern & NO other pattern	Special emphasis word "think".
			B	Possible Trend Patterns are removed from the historical data	Faster than Auto Model selection 1 & 2, as limited tests carried out.
			C	Autocorrelation test is carried out to check for seasonal pattern.	Vice-versa of strategy 51; Checks for Seasonal pattern.
			D	If no Seasonal pattern detected, system uses Constant model	
	53		A	Used if you **think** historical data has Seasonal and / or a Trend pattern	Uses strategy 51 and 52 or a seasonal trend model.
			B	Regression Analysis is run on Historical data to check for trend	Special emphasis word "think".
			C	Possible Trend Patterns are removed from the historical data	Faster than Auto Model selection 1 & 2, as limited tests carried out.
			D	Autocorrelation test is carried out to check for seasonal pattern.	Historical data does not exhibit constant pattern.
			E	If a Seasonal pattern is detected, system uses Seasonal model, If a Trend pattern is detected, system uses Trend model, If a Seasonal and Trend pattern is detected, system uses Seasonal Trend model.	
	54		A	Used if you **think** historical data has Trend pattern & you **know** that there is a Seasonal pattern.	Special emphasis word "think" & 'know'.
			B	Regression Analysis is run on Historical data to check for trend	This is a Manual + Auto model selection strategy.

			C	If Trend pattern detected, system uses Seasonal Trend model.	Vice-versa of strategy 55; Checks for Trend pattern.
			D	If no Trend pattern detected, system uses Seasonal model.	
	55		A	Used if you **think** historical data has Seasonal pattern & you **know** that there is a Trend pattern.	Special emphasis word "think" & 'know'.
			B	Autocorrelation is run on Historical data to check for Seasonality.	This is a Manual + Auto model selection strategy.
			C	If Seasonal pattern detected, system uses Seasonal Trend model.	Vice-versa of strategy 54; Checks for Seasonal pattern.
			D	If no Seasonal pattern detected, system uses Trend model.	
56 - Auto Selection 2			A	Used when no knowledge of pattern in historical data	Highly detailed tests carried out.
			B	Tests for Constant, Trend, Seasonal & Seasonal Trend.	More precise.
			C	Uses all possible combinations of Alpha, Beta & Gamma smoothing factors.	Longer time to run.
			D	Chooses model based on lowest MAD.	Not recommended for use in Mass Processing.

3.4.1.6 Croston Model

Forecast Method		**Forecast Strategy**		**Details**	**Observations**
80 - Croston Model			A	Used in case of sporadic or intermittent demand.	Generates one value as forecast for all forecasting

					periods.
			B		No Ex-post Forecast is calculated.
			C		None of the Errors calculated.
			D		Calculation based on # of data points, actual data value & intervals between data.
			E		Results differ based on where the data is located in the time horizon. Clustered data in the initial period or end of time horizon gives different results as also data well spread out in the entire time horizon.

3.4.1.7 Other Models

Forecast Method		**Forecast Strategy**		**Details**	**Observations**
60 - Historical Data			A	Used when demand does not change.	No forecast is calculated.
			B		Last year's history is copied to forecast.
70 - Manual Forecast			A	User manually specifies Basic value (Alpha), Trend value (Beta), Trend Dampening profile & Seasonal indices (Gamma)	Not to be used for mass processing.

				Parameters.	
			B		Not to be used for background jobs
			C		No Ex-post Forecast is calculated.
			D		Initially uses Forecast Strategy 40 to calculate forecast based on seasonal trend model.
94 - Linear Regression			A	Used when only single explanatory variable influences the forecasts.	Same principles as of MLR except only one independent variable is used.
			B		Based on least squares method.
			C		Relationship between the independent variable X and the dependent variable Y is assumed to be linear.
			D		Used most often for medium term (>3 & <= 24 months)
98 - No Forecast			A	No forecast done.	
99 - External Forecast			A	Forecast done by external systems.	User exits can be used.

3.4.2 Synopsis of time series models

In conclusion, selection of time series models can be summarized in the table below.

Pattern of data	**Model to select**
No Trend, No Seasonality	Moving Average Exponential Smoothing
Trend, No Seasonality	Holts method
Trend and Seasonality	Winters Model
No knowledge of patterns in historical data	Auto Selection 1 Think it is Trend Think it is Seasonal Think Seasonal/or Trend Think Trend, Know Seasonal Think Seasonal, Know Trend Auto Selection 2 System Selection
With Sporadic/ Intermittent Demand	Croston method

3.4.3 Causal Models (Multiple Linear Regression)

This method of forecasting is used to find the relationship between a single dependent variable and numerous independent variables.
Important requirements for MLR to be accurate in its forecast

1. There should not be a linear relationship BETWEEN the independent variable
2. The independent variables should not be stochastic – random events

The selection of the independent variables is one of the most important steps in a successful MLR forecast. APO uses the ordinary least square methods – which is one of the most popular methods of doing MLR

3.4.4 Time Series vs. Causal Model

In summary;

Time Series compares data being forecast over time. Here time is the independent variable (X variable)
Causal Models compare data being forecast against some other data set

Time Series methods assume that past trend and patterns will continue in the future.

Causal methods take into account what “caused” past data. Causal forecasting methods attempt to forecast future values of a time series variable (called the dependent variable) by using past data to estimate the relationship between the dependent variable and one or more independent variables.

Selecting between the above two models is determined by a variety of factors. One of them being the product's stage in it’s life cycle.

Product Life Cycle Stage	**Relevant Method**
New	Life Cycle [Pg 41] Opinion
Mature, unpromoted	Time Series
Mature, promoted	Time Series Cause/Effect
Seasonal	Life Cycle Time Series
Dependent	Causal
Cyclical (such as luxury items)	Causal

For example, forecasting sales of emerging products, which have little or no sales, history or a similar past product must rely on more qualitative techniques. As the product becomes more mature and more data is available, simple time series models become more useful. Causal models can ultimately be used with a rich data history.

Forecasting methods that can perform very effectively on good data may break down when applied to time series that are short or interrupted.

For example certain models require a minimum set of data and if these models are erroneously chosen, the software will go ahead and fit these models on request; however the results may ignore key patterns in the data and lead to implausible forecasts.

3.5 Statistical Tests

3.5.1 For Time Series Models

Statistical testing is conducted to help identify which of the forecasting methods chosen will work best with the available data.

It is to be noted that all forecasts will have errors. Two components make up the actual figures (sales) – systematic and random. The systematic component consists of Level, Trend, and Seasonality. The random events cannot be predicted. Hence the aim of the forecast method is to minimize the error in the systematic component.

Testing is necessary in order to avoid inappropriate methods and concentrate more effort on the methods that produce the most accurate forecasts. We can carry out testing to select the best method by producing ex post forecasts for past periods and comparing the forecast values with the actual values for those periods. When the best method has been found it is used to make ex ante forecasts.

Testing can also be used to refine the accuracy of the technique being employed since many methods have internal parameters, which can be varied [e.g. weights in weighted moving average, or the value of alpha in exponential smoothing].

Experiments and tests are used to find the most suitable parameter values. The overall objective is to minimize the error between forecast values and actual values. There are a number of different tests that are used for measuring forecast accuracy.

APO DP uses the following error analysis methods:

Test	**Measure**	**Comments**
Mean Absolute Deviation (MAD)	Average Absolute Deviation of forecast from actual.	The mean absolute deviation gives the mean average difference between the forecasted value and the historical value in the ex-post forecast. Best Practices: MAD to be as small as possible
Mean absolute percentage error (MAPE)	Absolute Error as a percentage of the forecast.	Same as MAD, except measures deviation as a percentage of actual data
Mean square error (MSE)	Variance of Forecast error	Not easily interpreted as MAD, MAPE. Not intuitive
Square root of the mean square error (RMSE)	Square root of Mean Square Error	Use instead of MAD.
Mean percentage error (MPE)	It is the difference between the history and the forecast; expressed as a percentage (divided by forecast) It is the same as MAPE	It is the same as MAPE.

	but without taking the absolute.	
Error total (ET)	It is the absolute value of the difference between the history and forecast value	It is the simplest of all errors.

In most cases the MAD and MAPE are used as the basis of comparison.

In the example shown below, two forecasting methods were used to generate errors. The errors from these two methods are listed below. Based on the customer's business process, the error that is deemed significant is used as the basis of the forecast model.

The MAD, ET, MSE and MAPE are calculated for two forecasting methods in the example below.

The results of this error show that the forecasting method 2 is more accurate based on the error methods chosen. It is possible that certain errors may be less in one method, while other errors are less in the alternate method. In such cases, the decision has to be based on the business requirements - the specific needs of the particular product/product group.

Forecasting Method1

Period	Forecast	Actual	MAD	ET	MSE	MAPE
1	500	480	6.0	20	400	4.2%
2	550	525	11.7	25	625	4.8%
3	575	530	21.7	45	2025	8.5%
4	625	580	28.7	45	2025	7.8%
5	680	720	32.0	40	1600	5.6%
Average			32.0	35	1335	6.1%

Forecasting Method 2

Period	Forecast	Actual	MAD	MAE	MSE	MPE
1	470	480	3.0	10	10	2.1%
2	530	525	3.6	5	25	1.0%
3	560	530	11.5	30	900	5.7%
4	600	580	14.0	20	400	3.4%
5	715	720	11.3	5	25	0.7%
Average			11.3	14	290	2.6%

3.5.2 For MLR Models

After the MLR is selected, the forecast results are tested using one or more of the following tests

Test	Measure	Comments
R Square	Goodness –of-fit	0 = Model not explained 1 = Model perfect fit Explains how the chosen X's the Y
Adjusted R	Proportion of	If this is significantly lower than R Square, an X is

Square	Variation in Y	missing Use the same X when comparing two models
Durbin H	Autocorrelation [Pg 42] in time series (If Value >=1.96)	X's lag by one or more periods Can be used only if there are more than 100 time series values.
Durbin Watson	First order autocorrelation (Acceptable range1.5 to 2.5)	No lag between the time series values
T-Test	Is a particular X to be used in the model	SAP recommends +/-1.4 to be left in the model
Elasticity	Measures the effect of 1% change in Y	Unit Free Higher the number, more responsive Y is to changes in X

4. Reasons for Ineffective Forecasting

- Not involving a broad cross section of people
- Not recognizing that forecasting is integral to business planning
- Not recognizing that forecasts will always be wrong
- Not forecasting right things
- Not selecting an appropriate forecasting method
- Not tracking accuracy of forecasting models

5. Conclusion

In summary, after the Data analysis has been completed, the appropriate model has been selected & the forecast executed, then the forecast is ready for release to other systems.

For Example:

- EXCEL
- OLTP
- BW
- LEGACY
- OTHER

6. Appendix

6.1 Useful Web Sites

International Institute of Forecasting: http://forecasting.cwru.edu/
Principles of Forecasting (Wharton): http://morris.wharton.upenn.edu/forecast/
The Institute for Forecasting Education: http://www.forecastingeducation.com/

Other links

http://www-marketing.wharton.upenn.edu/forecast
http://www.tai.hut.fi/ecomlog/publications/use.html
http://www.hacketthighway.com/bpc_bp_qw.asp?PgId=32
http://www.infochain.org/quarterly/feb97/lq2_pg23.html
http://www.itl.nist.gov/div898
http://www.cosc.brocku.ca/~duentsch/archive/stattech.pdf
http://www.statsoftinc.com/textbook
http://obelia.jde.aca.mmu.ac.uk/resdesgn
http://ubmail.ubalt.edu/~harsham/Business-stat

6.2 Books & Articles

Forecasting: Methods and Applications. By Spyros G. Makridakis, Steven C. Wheelwright (Editor), Rob J Hyndman

6.3 Details / Best Practices for Workshops

< 1 >

Details / Best Practice	Workshop
The products that are critical to the business should be forecasted. For example criticality, could be based either on any one or more of the following criteria **(A) Profitability:** The contribution of this product or group of products to the company overall profit margins. **(B) Product Life Cycle [Pg 41]** **(C) Influence of Promotions** **(D) Volume**	Data analysis

(E) Lead Time **(F) Resource Constraints**	

< 2 >

Details / Best Practice	Workshop
Data can be stored in days but rather forecast in weeks or months. Typically we analyse data at the lowest level, depending on the nature of the business. We then decide the level of aggregation to be used for forecasting ex: weekly vs Monthly. (Typically we don't recommend data to be run at the daily level for long term forecasting, for reasons of sizing and performance. Also the forecast is typically not used at the daily level)	Data analysis

< 3 >

Details / Best Practice	Workshop
This depends on the nature of the business and the specific business requirements. Business requirements drive the frequency of forecasting.	Model Selection

< 4 >

Details / Best Practice	Workshop
This depends on a variety of factors, some of them include: (A) Specific business requirements. Although there is no single correct answer, to reliably fit and test statistical models, a time series of 48-60 months is desirable. At least 3 seasonal cycles are in order for estimating a seasonal model. With 48 months, for example, the first 36 can be used to fit the model and the last 12 to test the model's forecasting accuracy. On the other hand, for short-term forecasting, going back more than 60 months is unlikely to be helpful, since most statistical methods assign more weight to the recent than to the distant past. Use the largest sample possible. (B) Data availability is one of the Factors Influencing Model Selection. [Pg 14]	Model Selection

< 5 >

Details / Best Practice	Workshop
The basic premise of time series forecasting is that “history repeats itself”. But in case of major structural or economic changes, the historical data needs to reflect the changes. “Corrected History” needs to be used which corrects the outliers, highlight past promotions if any & also to reflect mergers or acquisitions. Identify patterns from existing historical data.	Data analysis

< 6 >

Details / Best Practice	Workshop
If a new product is being introduced, history can be created using like modelling. Create history for like (similar) products using like modelling.	Model Selection

< 7 >

Details / Best Practice	Workshop
Data patterns for certain products can be easily highlighted through plotting (Data Analysis). Though for certain products, the pattern is not clearly visible. Hence use of techniques, which facilitate automatic model selection. These methods verify various parameters simultaneously to find the best-fit model.	Model Selection

< 8 >

Details / Best Practice	Workshop
After data analysis, patterns should emerge that enable model selection. Analyse data to establish patterns.	Model Selection

< 9 >

Details / Best Practice	Workshop
Products with intermittent demand cannot be treated similar to products with regular demand. The frequency of the demand pattern is critical. These products need to be clubbed together.	Data analysis/Model Selection

< 10 >

Details / Best Practice	Workshop
Since outliers influence the results of almost all statistical estimators and testing procedures you must remove outliers from your data.	Data analysis

< 11 >

Details / Best Practice	Workshop
The number of workdays impacts the forecasted quantities. Hence it needs to be taken into consideration.	Model selection

< 12 >

Details / Best Practice	Workshop
If the customer is aware of the impact of Causal factors and has data to support the assumptions, then they need to be collected and loaded into the application	Data analysis

< 13 >

Details / Best Practice	Workshop
If so, recognize the difference between planning values generated by different models. Do not compare the forecasts generated by different models for different products.	Model Selection

< 14 >

Details / Best Practice	Workshop
If two or more methods are combined, the weightage of each of these methods must be clearly understood.	Model Selection

< 15 >

Details / Best Practice	Workshop
If past promotions exist, they can be used as a benchmark for any future promotions. Past promotions can be copied and edited to facilitate the creation of new promotions.	(Applicable to both model selection and data analysis)

< 16 >

Details / Best Practice	Workshop
This is typically the role of a statistician. (In some cases the data analysis is not required, since the Demand Planner is experienced enough to suggest a forecast model.) If the	Data analysis

statistician is not available this process is to be done jointly by the Demand Planner and the consultant.	

< 17>

Details / Best Practice	Workshop
The 80-20 rule could be applicable in many cases where20% of the products contributes to 80% of the profits. Fast vs. Slow moving products Typically 80% of the products maybe able to use forecast method A and 20% could use forecast method B. Another scenario would be the slow movers could be manually forecasted vs forecasting fast movers using a selected model.	Model Selection

< 18 >

Details / Best Practice	Workshop
Forecasting aggregated entities typically results in lower forecast errors than forecasting individual entities. This is due to the principal of compensating errors, in which errors cancel each other out. For example, forecasts of product groups are more accurate than forecasts for each SKU in the group, and forecasting for several periods of time is more accurate than forecasting for an individual period.	Model Selection

< 19 >

Details / Best Practice	Workshop
Reliable quantitative historical data has to be available. Time series methods would not be appropriate where relevant historical data is not available. Nor would they be appropriate where the historical data is unreliable, perhaps because it has not been recorded in a consistent manner over the relevant time period. One has to be certain that information is both reliable and in a usable form. If the data has	Model Selection

low reliability then you will need to increase the sample size.	

< 20 >

Details / Best Practice	Workshop
Different forecast accuracy measures are available. Depending on the business requirements a specific measure is chosen and that value is used to select the model.	Model Selection

< 21 >

Details / Best Practice	Workshop
This is generally not recommended as it involves development efforts & budgets. However, unique products & situations / industries may require this. This can be handled by USER EXITS.	Model Selection

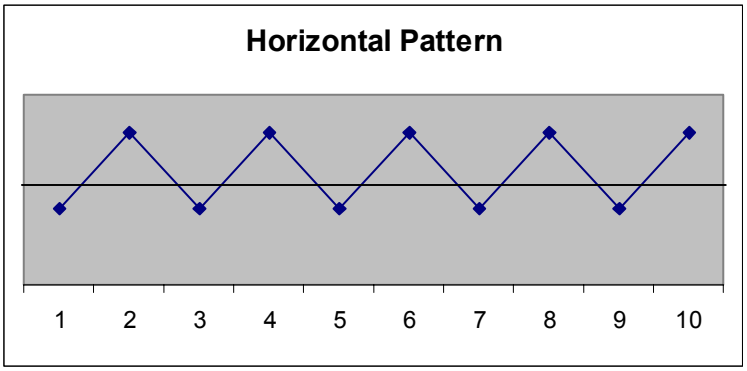
< 22 >

Details / Best Practice	Workshop
In situations, where the planning of the dependant demand is critical as the main product, this methodology is recommended. APO supports SINGLE INDENTURE BOMs.	(Applicable to both model selection and data analysis)

6.4 Statistical Patterns

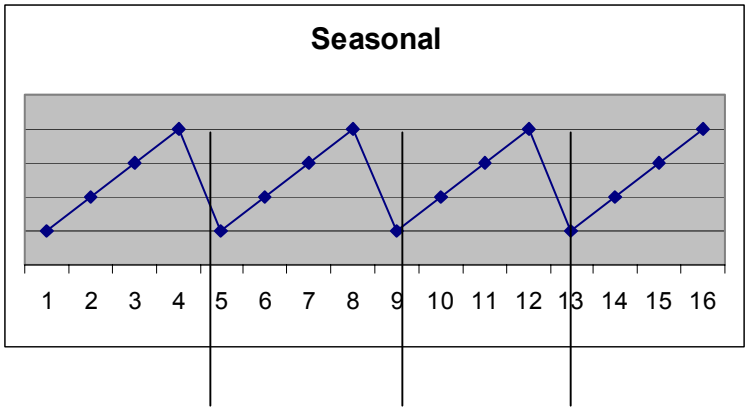
Horizontal Pattern

No trend, stationary. Equally likely chance that the next value will be above or below the mean. Ex; Stable sales.



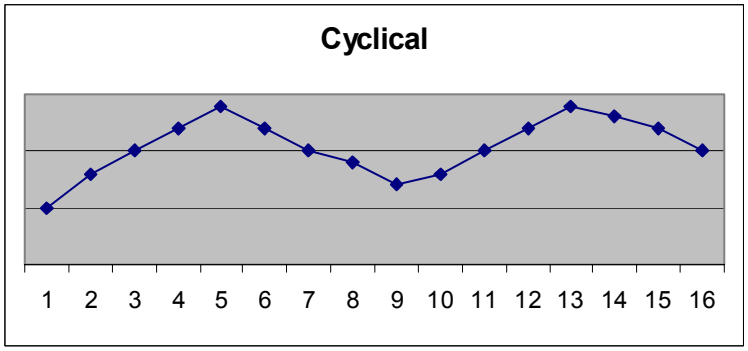
Seasonal Pattern

Periodic Patterns in a time series that complete themselves within a calendar year. The pattern also repeats itself each year.
It is caused by:
Production/harvesting patterns
Weather
Customs

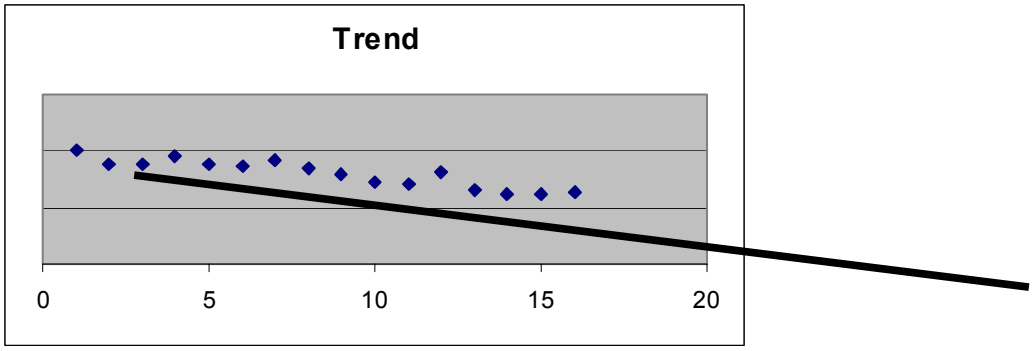


Cyclical Pattern

This refers to **recurring** up and down movement **around** TREND levels. Cycles can last 2 to 10 years in length. Cycles can be caused by changes in taste and preferences due to the whims of fashion. They vary in length and magnitude from one cycle to another.



Trend Pattern



A general upward or downward movement of a series over some period of time. Trend represents long-term growth or decay. They usually have a strong underlying explanation such as:
Technological change
Changes in consumer preferences & tastes
Changes in per capita income
Changes in the number of consumers
Market competition
Inflation or deflation
Policy change

Random or Irregular Pattern

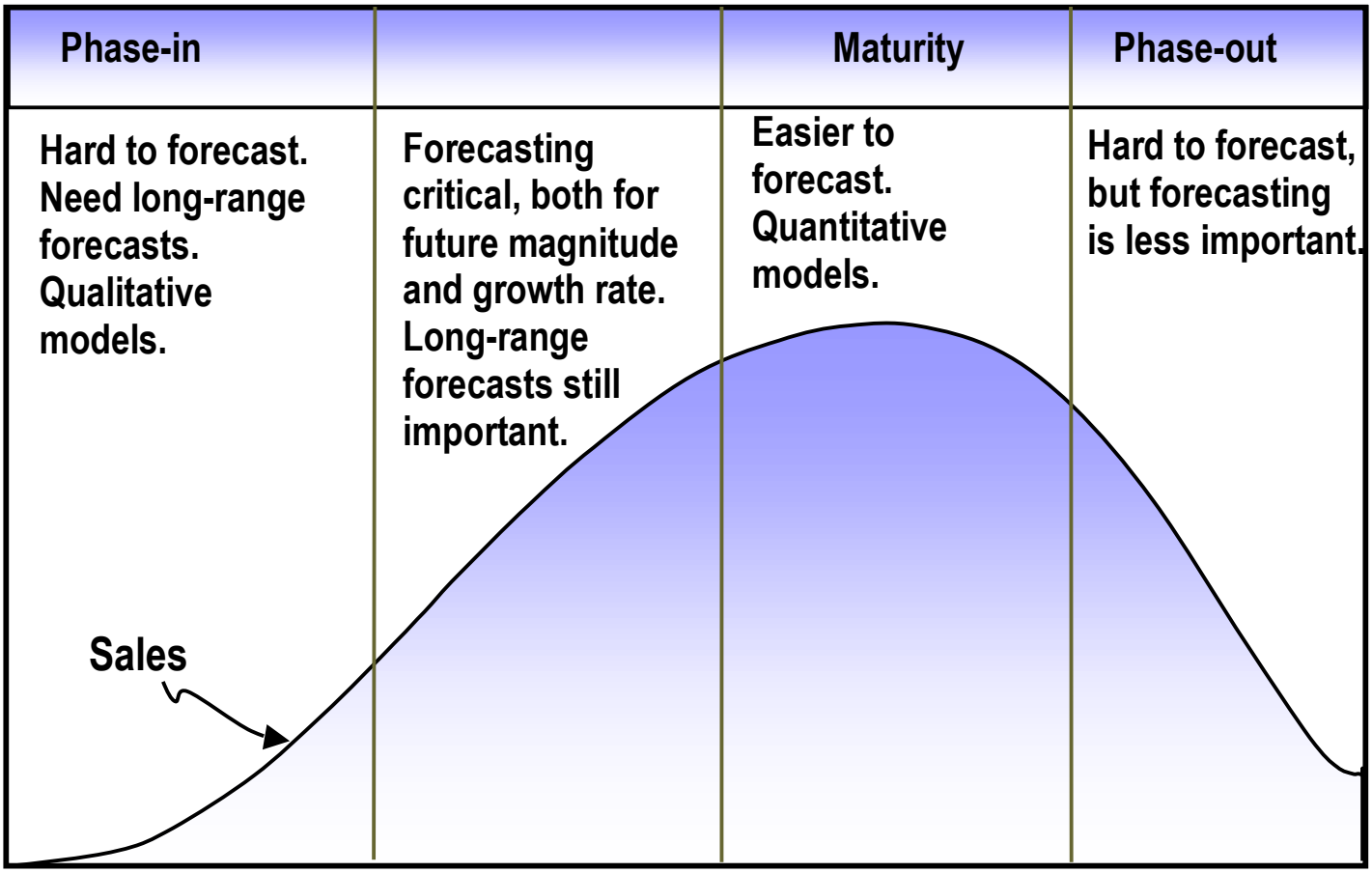
Erratic, Nonsystematic, Random, 'Residual' Fluctuations.
Due to random variations of:

- Nature
- Accidents

Short Duration & Non - Repeating

6.5 Product Life Cycle

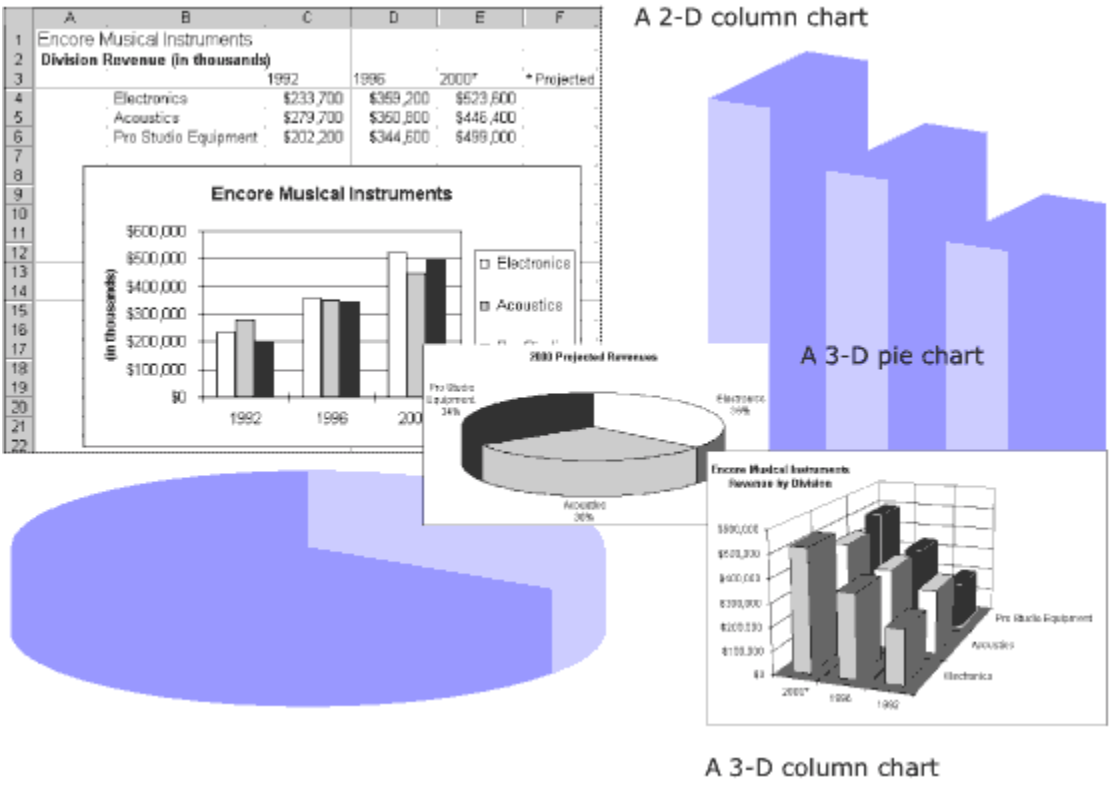
The 4 different stages in a Life Cycle of the product are as below:



6.6 Acronym Guide

Alpha:	Smoothing factor for the Basic value in a Constant Model
Beta:	Trend factor for the trend value in a Trend Model
Gamma:	Seasonal factor for the Seasonal Indices in a Seasonal or Seasonal/Trend Model
Autocorrelation:	Is a tool for identifying dependence of current data on previous data points. It is undesirable since it can falsify the demand forecast.
BW:	Business Warehouse
ET :	Error Total
Ex-Post Forecast:	This is the forecast created by the system, for the past periods, using the alpha, beta, gamma values that was generated, based on the historical data
Legacy System:	Any old computer system, which is home grown/Main Frame
MAD:	Mean Average Deviation
MAPE:	Mean Absolute Percentage Error
MPE:	Mean Percentage Error
MSE:	Mean Square Error
MLR:	Multiple Linear Regressions
OLTP:	On-Line Transaction Processing
OLAP:	On-Line Analytical Processing
PERMSO:	Periods in Smoothing factor
RMSE:	Square root of Mean Square Error

6.7 Analysing the data - Links

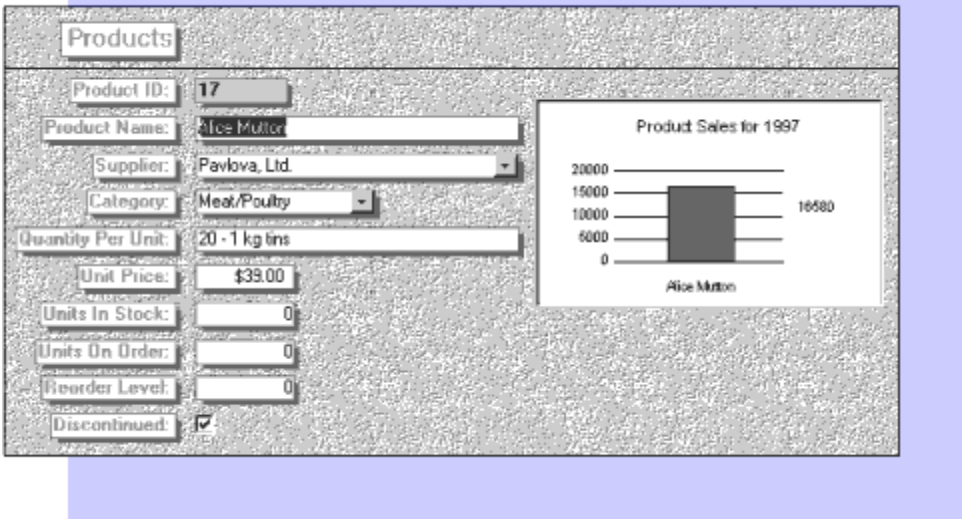


For data stored in Microsoft Excel, you can display it graphically in a chart. Show the values as lines, bars, columns, pie slices, and other data markers, and even combine different markers in the same chart.

If the chart is all that is needed, a separate chart sheet can be created. When it's best to display the chart along with the associated data, you can either create a chart directly on the worksheet or add a table to a chart sheet containing the pertinent data. In any case, the values in the chart are updated whenever the source worksheet data changes.

For data stored in Microsoft Access, you can represent your data in chart form and then add the chart to a database form or report. To create a chart quickly, you can use the Chart Wizard.

When you browse through the Products form, the chart changes with each record to display information about the currently selected product.



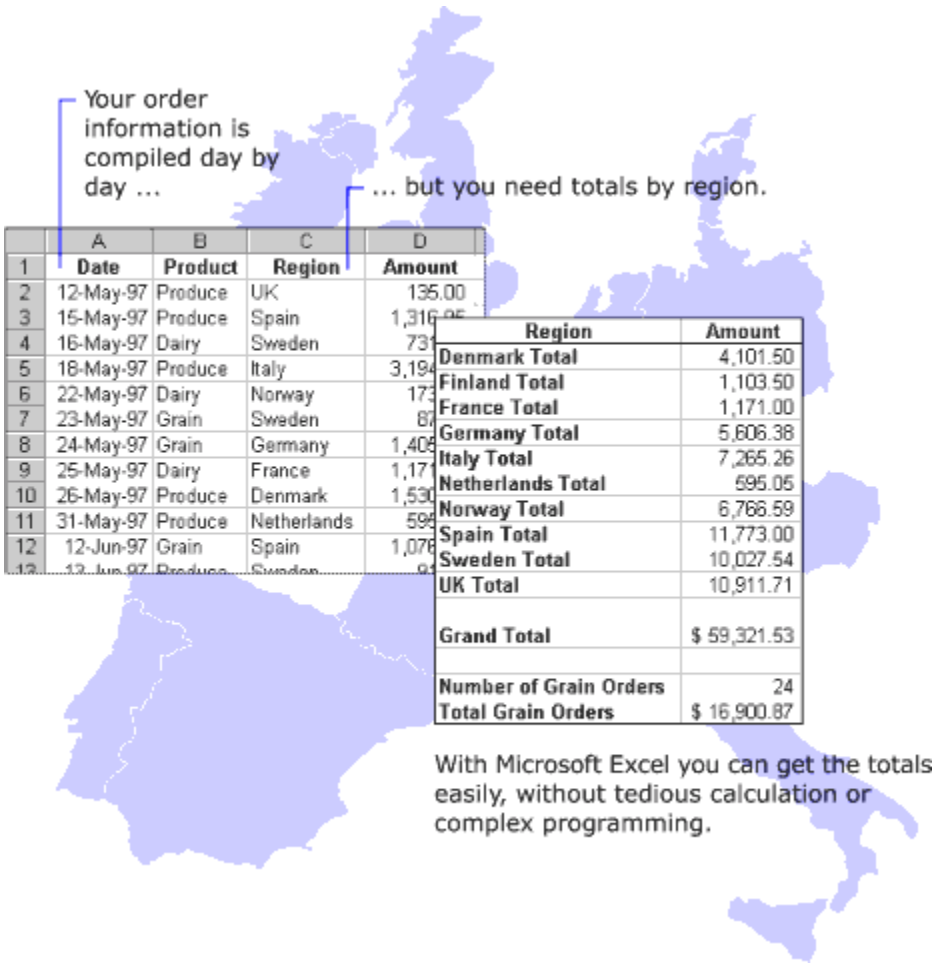
For Data Stored in Microsoft Excel

If you want to	Do this
Create a detailed report that organizes, subtotals, and summarizes your data	Add automatic subtotals to your data, Ex: Create a Detailed Sales Report.
Create a chart that summarizes your data graphically	Run the Microsoft Excel Chart Wizard.
Create a summary table that lets you change your view of the data dynamically	Create a Microsoft Excel PivotTable.
Create a detailed Microsoft Access report without making changes to your original worksheet	Run the Microsoft Access Report Wizard directly from your Microsoft Excel worksheet. Click **MS Access Report** (**Data** menu). This command appears only if the Access Links add-in is installed and available. Click **Add-Ins** (**Tools** menu), and then click **Access Links Add-In**. If this option does not

appear, rerun setup and install the Data Access options.

Ex: **Creating a detailed sales report**

Do you have detailed data and want to see totals? For example, suppose you receive information about orders as each is filled over the course of several months. You might need to calculate the total sales for each region and the total product sales across the regions. Microsoft Excel can rapidly organize and sum up this kind of data for you.



	A	B	C	D
1	Date	Product	Region	Amount
2	12-May-97	Produce	UK	135.00
3	15-May-97	Produce	Spain	1,316
4	16-May-97	Dairy	Sweden	73
5	18-May-97	Produce	Italy	3,19
6	22-May-97	Dairy	Norway	17
7	23-May-97	Grain	Sweden	8
8	24-May-97	Grain	Germany	1,40
9	25-May-97	Dairy	France	1,17
10	26-May-97	Produce	Denmark	1,53
11	31-May-97	Produce	Netherlands	59
12	12-Jun-97	Grain	Spain	1,07

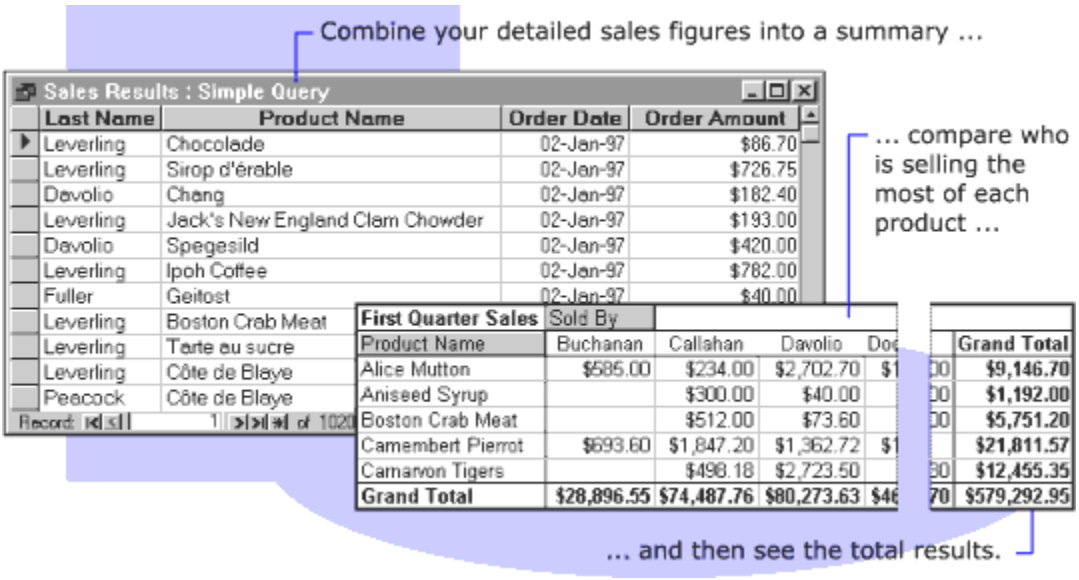
Region	Amount
Denmark Total	4,101.50
Finland Total	1,103.50
France Total	1,171.00
Germany Total	5,606.38
Italy Total	7,265.26
Netherlands Total	595.05
Norway Total	6,766.59
Spain Total	11,773.00
Sweden Total	10,027.54
UK Total	10,911.71
Grand Total	$ 59,321.53
Number of Grain Orders	24
Total Grain Orders	$ 16,900.87

For Data Stored in Microsoft Access

Use the following table to decide which topics to read for more information.

If you want to	Do this
Create a detailed report that organizes, subtotals, and summarizes your data	Run the Microsoft Access Report Wizard.
Create a chart that summarizes your data graphically	Run the Microsoft Access Chart Wizard.
Create a Microsoft Excel summary table that lets you change your view of the data dynamically	Create a Microsoft Excel PivotTable.
Organize, subtotal, and summarize the data by using Microsoft Excel	In Microsoft Access, click **Office Links** (**Tools** menu), and then click **Analyse It With MS Excel** to export a snapshot of the data to a Microsoft Excel worksheet. Create a Detailed Sales Report. When your data changes, you must repeat these steps for updated results.

Create a Sales Summary from a Microsoft Access Database



Use Office Applications Together

If Your Data Is Stored Somewhere Other Than in Microsoft Access or Microsoft Excel

Both Microsoft Access and Microsoft Excel let you work with data from external sources.

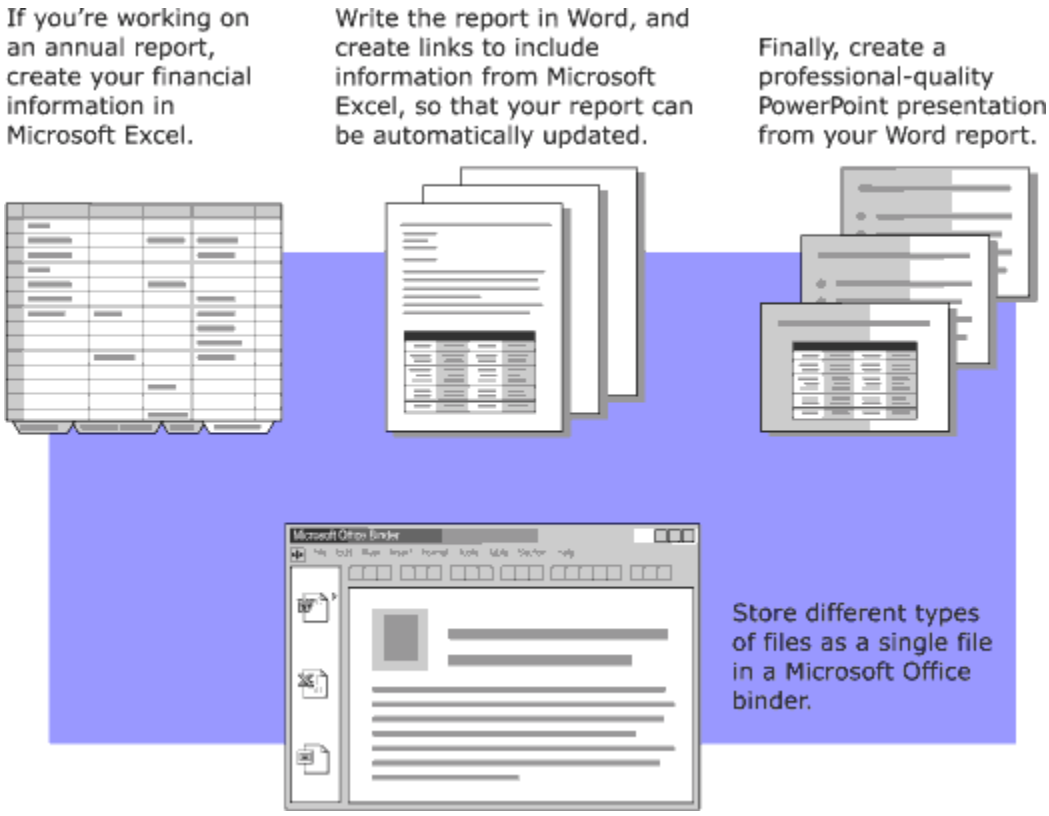
Import files into or link files to your Microsoft Access database Importing a file copies a snapshot of its contents into your database. Creating a link allows you to work with a file that continues to be maintained in its originating application. Use Office Applications Together.

Bring data into your Microsoft Excel worksheet by using Microsoft Query You can analyse external data in Microsoft Excel and refresh the data in your worksheet when it changes. Get Sales Information from a Database.

Use Office Applications Together

It's easy to use Office applications together. For example, you can create a Microsoft PowerPoint presentation from a Microsoft Word outline. Probably the most common way to share information between applications is to copy and paste information from one application to another. You can also create a **link** from one application to another or insert objects to include information. In online documents, you can use **hyperlinks** to jump to other files on your organization's internal Web (**intranet**) or sites on the World Wide Web.

Combine information from different applications by using the Microsoft Office Binder to organize and print files with continuous page numbers from Microsoft Excel, Word, and PowerPoint as a single unit.



Get Sales Information from a Database

Chances are you sometimes work with data that isn't stored on your computer. For example, your company may compile sales statistics in a database located on a networked mainframe.

Out of mountains of detailed data in the corporate database, extract just what is relevant to your work by using a **query**—a method of extracting specific data from a database. Then you can use familiar spreadsheet analysis tools on that data, without retyping it.

Select just the data you need, and return it to Microsoft Excel for further analysis.