FORECASTING WITH TIME SERIES ANALYSIS

by

Graham Dobie Armstrong

submitted in fulfilment of the
requirements for the degree of

MASTER OF SOCIAL SCIENCE

in the Department of Business Administration,
University of Natal, Durban

Supervisor: John MacDonald, D.Com.

Co-supervisor: Adolf Diegel, Ph.D.

January 1998

SUMMARY

This thesis was undertaken with the intention of applying forecasting with time series analysis, in a manufacturing context. This involved two phases: the updating of existing forecasting techniques, and the application of these techniques to a manufacturing firm.

The existing techniques, developed mainly by Brown in the 1960's, had to be adapted for computer application, to allow fast and objective computation of forecasts. This required an investigation into the derivation of each algebraic model, previously computed by hand, and translating those intuitive steps into routine ones. Furthermore, the revision of each forecast in the light of new data had to be dealt with mechanically.

As for the application, the data supplied by the client, a large South African manufacturing firm, did not permit a successful application. This concerned both the manner in which the data were recorded (inconsistent time intervals), and the volume of data readily accessible. This then led the thesis in an unanticipated direction to overcome these difficulties. To do this objectively, it became necessary to generate test data with known characteristics, then to study how many data were required to recover those characteristics.

Generating data required an investigation into random number generation, real data consisting of both true changes as well as a percentage of random fluctuations. A random data series was, therefore, added to the series with known characteristics. Such characteristics are unknown for genuine data, such as those supplied by the client. Empirical experimentation with the generated data, led to the determination of the number of data required to recover coefficients of various complexity. This number was found to be contrary to the statements made by Brown on this topic, significantly more data being required than was previously thought.

Finally, an attempt was made to select an appropriate model for the client's data, based on the knowledge gained from investigating generated data.

DECLARATION

This thesis, unless specifically indicated to the contrary in the text, is my own original work. It has not been submitted for examination to, nor am I registered with, any university other than the University of Natal, Durban.

Graham Armstrong

LIST OF FIGURES

LIST OF TABLES

Table	*Page*

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

Adolf Diegel, without whose encyclopaedic knowledge and tireless commitment, this thesis would not have been possible.

John MacDonald, as supervisor, for his assistance in putting the thesis together.

Glen Polly, planning manager at Lever Ponds, and

Bala Chetty, forecasting co-ordinator at Lever Ponds, for making their data available.

CHAPTER 1

What we want and what we know

1.1 Introduction

Most people, especially businessmen, would like to be able to anticipate future events, such as demand and prices, yet few can actually see such events beforehand, by literally looking into the future. All we can see is past data.

Sometimes events which happen today affect those in the future. For example, a change in the gold price affects the flow of money into South Africa, some time later: there seems to be a definite *causal* relationship between one and the other. More precisely there is a *lagged* causal relationship, the effect taking place so many months after the cause, for example, three months. This is useful enough if such an horizon is adequate. But what if one were interested in events four or six months ahead? One would first have to know the gold price over the next few months, while projecting for a lag of three months. As a rule, it is difficult, if not impossible, to anticipate the occurrence of the next causal event any better than the next effect. Will there be good rain next year, or maybe a flood or drought? How will the gold price fluctuate in the future?

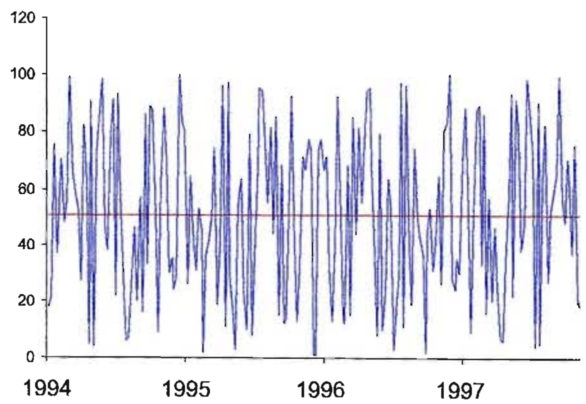
In answering these questions, one eventually needs to resort to the same methods to forecast presumed *causes*, as one does to forecast *effects*. So one might as well deal with the effects directly.

1.2 Forecast models

We can *know* only the past. The past is reflected in the data kept by companies. What one is looking for in these past data are patterns and repetitive processes such as trends and seasons. Once these are found to have persisted in the past, it is expected that they will continue to repeat themselves in the future. Such patterns are represented by more and more sophisticated mathematical models. Typical models relevant to commercial data are discussed in the following paragraphs.

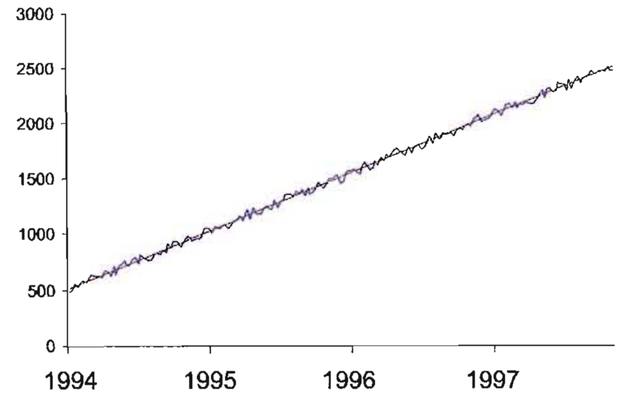
Often data fluctuate strongly over time. Figure 1.2.1 illustrates such fluctuations in raw data over a four year period, shown in blue. The variations in the data, at face value, are not regular. In other words, they go up and down without any obvious pattern. In such a case, the underlying process could be described as constant, shown in red in Figure 1.2.1. The oscillations can be more or less pronounced, but to the extent that they are random, one may hardly hope to discover a pattern in them.

Figure 1.2.1: Constant model, with seemingly random fluctuations



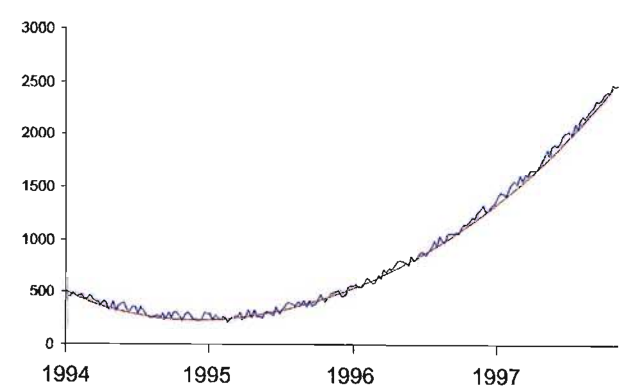
Equally often, sales data are regularly increasing (or decreasing) over time, as shown in blue in Figure 1.2.2. In such a case one may want to study the *rate of change* in the data over time. To the extent that it is regular, one calls this a *trend line* or *linear model*.

Figure 1.2.2: Linear model



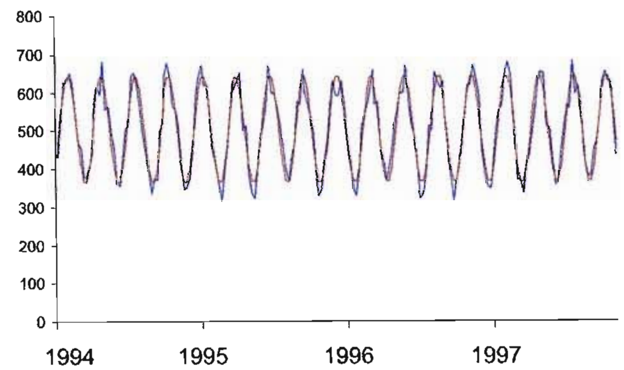
Another form of process that may characterise sales data is where the sales may change very rapidly over time. For example, this can happen when a new product is introduced to the market, again the raw data are shown in blue in Figure 1.2.3. In these data, the rate of change is itself changing, becoming more and more pronounced as the product catches on. The question then concerns the *rate of change of the rate of change*. It could itself be constant, as shown in red in Figure 1.2.3, or the rate of change may itself continue to change. Finally, that change in the rate of change may eventually turn out to be a percentage change (exponential model).

Figure 1.2.3: Curve model



Still another model often found in sales data, is the *seasonal model*. As its name implies, it tends to reflect the seasons of the year, or some other repetitive process. One then uses a trigonometric function (sine and cosine) to model the data. This is shown in Figure 1.2.4, where the blue line represents the raw data, and the red line, the seasonal model.

Figure 1.2.4: Seasonal model



The previous graphs represent fundamental models, but in reality, they may be found as such or in any combination, for example, of a trend and a season.

Just as models can be shown graphically, so they can be represented by a formula or equation. The task is to find such a formula for a given set of data, and to consider which model *best* fits them. Generally speaking, that model is best which not only fits the past data well, but also needs little revision as the future unfolds. If it needs very little revision, it can be said to correctly describe the "true" underlying process, and thus be able to anticipate the behaviour of future data with similar precision. However, it is unusual to find a model which anticipates the behaviour of data precisely. It then becomes necessary to ensure that the forecasts produced are kept up to date.

1.3 Keeping forecasts up to date

In revising the model from day to day, *new* data tend to be more important than older ones, as they reflect what is currently happening in the market, rather than yesterday or last month. One may, therefore, want to give recent data comparatively more weight than past data, so as to keep the model up to date with the unfolding process.

On the other hand, if a forecast requires *rapid* updating and revision to remain accurate, the model itself may require revision, the *best* model being the one that requires the *least* revision. The errors or mistakes made in forecasting, for example, the difference between last month's forecast and last month's datum, also should be considered in future forecasts. The good forecast thus tends to include a small amount of recent data, and the accumulated effects of previous forecasts. However, the *goodness* of a forecast is also dependent on the quality and quantity of the data recorded.

1.4 Applying time series analysis to manufacturing

A crucial aspect of data recorded over time, is the length of the interval between observations. This interval can be practically nil (*continuous* time series), or of measurable duration (*discrete* time series). In the latter case, the interval can be *regular* or *irregular*.

Business data in particular, tend to be observed at irregular intervals. Even if the time unit is a seemingly constant unit, such as a week, some weeks contain fewer sales days than others. This is even more seriously the case for monthly data, largely the norm in manufacturing firms. Not only are months of unequal lengths in the calendar (between 28 and 31 days), but also one has various holidays, weekends, strikes and other days without work.

Another crucial aspect of forecasting in a manufacturing context, is the distinction between *sales* and *demand* data. The two may differ significantly if stock was insufficient to meet current demand. Obviously, forecasting should be based on demand so that a manufacturer can capture the maximum number of sales. In fact, it may be difficult to observe true demand, because unfilled demand tends not to be recorded. Demand data would have to be collected at the point of sale, that is, where the consumer asks for goods, rather than from accumulated orders from distribution warehouses. In the case of a large manufacturing firm, these figures would have to be sourced from the outlets which they supply. This would involve close cooperation and the synchronisation of database systems.

One can deal with the irregular time intervals by knowing the true gap between successive observations. For example, monthly or weekly data need to be recorded on the last day of each respective interval, so that the observations are a reflection of that full time period. The problem is reduced if observations are made on a daily basis, as shopping days tend to be of more equal durations. As regards the demand *versus* sales data aspect, one can only be aware of it in assessing the quality of the forecast, but it is difficult to deal with it quantitatively.

1.5 The quality of data and the forecast

This thesis was commenced with the intention of applying standard forecasting procedures specifically to a South African manufacturing firm. However, in *testing* or *developing* a forecasting program, a number of difficulties need to be overcome. These can be grouped into two dimensions: a general and a specific dimension. The *general* dimension contains the problems related to forecasting as a discipline, namely:

- substantiating objective criteria for the selection of an appropriate model to correctly describe the underlying process;
- the degree to which the model chosen requires revision in the light of new data;
- adapting the model selection and revision procedures to be performed by computers.

The possibly enormous volume of data available, both in the number of products and archived records, necessitates the intervention of computers because of their speed and accuracy of computation: after all, one should like a forecast to be ready *before* the event in question occurs (real time forecasting). Computers also provide the objectivity that makes routine forecasting possible. In turn, this requires forecasting procedures based exclusively on data rather than opinions, and on mechanical/objective computation, untouched by human hands.

The *specific* dimension refers to problems related to the way data are recorded for manufacturing processes. It includes the following:

- the distinction between demand and sales data;
- the irregularity, disorganization and quantity of the data evaluated;
- the inability of these data to reveal their underlying process.

As concerns the quality of the data, they could turn out to be surprisingly disorganised and insufficient. In such a scenario, even if more complete data were available, they may still not show which model they reflect. To appreciate this point, one must take account of the fact that real data usually contain a significant random element. The question then arises how to separate that randomness from the underlying process. Clearly, it would emerge if one looked at "many" data, but how many is enough? That question cannot be answered experimentally by working with real data, since one does not know which underlying process they represent.

However, if one *generates*, for example, a trend line or seasonal data with known characteristics and adds a given random element, one then knows beforehand what one is dealing with. In turn, one can study how many observations must be made to recover those known elements. If a forecasting program is successful with generated data, with a degree of randomness, then it can be trusted to be successful with real data. Therefore, the development of a successful forecasting model will entail working with artificial/generated data as much as with real ones.

Successful forecasting requires a good quality model, but this depends on having good quality data from which to establish the model. The success of the attempted application of forecasting to a manufacturing firm, depends on how well these difficulties can be overcome or dealt with.

1.6 Approach to the thesis

Just as the problems identified can be classed into two categories, so this thesis will consist of two parts. In dealing with the general problems identified, chapters two to six will include a discussion of established forecasting techniques, as well as their adaptation to a computer. Chapters seven to eleven will include the methodology for dealing with the specific problems encountered, and the attempted application to manufacturing data. Chapter twelve provides recommendations and the conclusion that can be drawn from the findings in the thesis.

1.7 Summary

This thesis will involve using the power of a modern computer to test established forecasting techniques against generated data. It will then be attempted to apply these techniques to a South African manufacturing firm.

CHAPTER 10

Varying the random element

10.1 Introduction

In the previous chapter, the formula for computing the number of data required by a given model, was derived from two different random series, each of a constant size. The same random element was used with different models. The exact exponent and multiplier for computing the number of data required to successfully retrieve the known coefficients ($2^{time} \times 15$), was therefore derived from these particular sets of random numbers.

The typical randomness of *real* observations is not fixed, but depends strongly on a particular type of product. The random data used varied by ±50, that is, they ranged from 1 to 100. It would seem reasonable to expect that a relatively larger random variation in a given set of data, would make the true coefficients more difficult to recover. Similarly, by reducing this variation, one would expect to recover them sooner.

10.2 The size of the random element

Focussing on the trend, it was shown that the true coefficients can be retrieved from 30 observations. Changing the origin or the strength of the slope, or both, would have the effect of enlarging or reducing the *relative* percentage of randomness in that data. However, increasing or decreasing the size of the random numbers themselves, say by doubling or halving, will have the same effect, although the new size of the random numbers will be known. Simple experimentation can again demonstrate the effect this would have on coefficient recovery, knowing the true coefficients and the size of the random factor.

The data used to test the changed random content can be seen in Appendix 7. These sets contain different magnitudes of the same generated random numbers, added to the same simple trend previously used (origin 450 and slope 10). The random element added to the trend was the same as in the previous chapter (modulator 18, multiplier 101), but doubled, kept in its original form, divided by two, and again by two. Any fractions were rounded to the nearest whole number. There are thus four sets of data for the trend, with random variations of ±100, ±50, ±25 and ±13.

10.3 Retrieving the coefficients

The same fitting procedure as in the previous chapter was used. That is, a group of 15, 30 and 60 data were used to fit the coefficients, from three different starting points in the data, and then forecast for the next 50 weeks. The residual variance for each group was compared to the residual variance for using 100 data in fitting, the so-called best fit. The series were not tested in reverse order as before, because one obtains the same results whether fitting in normal or reverse. The results of fitting the trend data with a varied random element, are shown in Table 10.3.1.

Astonishingly, the change in the quality of fit with 30 data (shaded in Table 10.3.1), while increasing and decreasing the size of the random element, was negligible. Indeed, the results were practically identical for all fits, because the individual results remained in the same proportion, no matter what the relative size of the random element.

Table 10.3.1: The results of fitting trend data with different degrees of randomness

Group (Start)	Trend (450 + 10 × time)							
	100 variation		50 variation		25 variation		13 variation	
	Sigma	%	Sigma	%	Sigma	%	Sigma	%
All	57.7444	100	28.8722	100	14.4334	100	7.1984	100
Small	15 data							
(1)	159.4153	276	79.7077	276	39.6045	274	20.5470	285
(21)	256.5852	444	128.2926	444	63.9842	443	33.0586	459
(41)	85.0654	147	42.5327	147	20.6810	143	10.7832	150
Average	167.0220	289	83.5110	289	41.4232	287	21.4629	298
Medium	30 data							
(1)	65.5804	113	31.7902	110	15.7359	109	8.2019	114
(21)	58.9094	102	29.4547	102	14.7150	102	7.4885	104
(41)	68.5348	119	34.2674	119	17.0468	118	8.5826	119
Average	64.3415	111	31.8374	110	15.8326	110	8.8091	112
Large	60 data							
(1)	61.4900	107	30.7450	106	15.3550	106	7.7193	107
(21)	63.3713	110	31.6857	110	15.7863	109	7.9364	110
(41)	57.9458	100	28.9729	100	14.4840	100	7.2175	100
Average	60.9357	106	30.4679	106	15.2084	105	7.6244	106

Tests with reducing the slope, origin or both, while keeping the same random series, had the same negligible affect. Furthermore, random variations as low as ±3, still substantiated the results shown in Table 10.3.1. No matter how the random element was changed, large or small, 30 data were still required to successfully recover the trend coefficients. This may seem contrary to what was expected initially, but upon reflection, it stands to reason: the amount of data required to recover a model's coefficients depends on the model's responsiveness, not on the size of the random element it responds to. In other words, a model as responsive as a trend will be misled by small variations just as easily as by large ones.

The initial generated random series used in fitting was generated with a multiplier of 18. It was demonstrated that this series was the "most purely random" of those generated with the modulator

101 and seed of 1, as shown in Table 8.4.1. But what would the effect be of using a slightly less random series: less random in that it contains pockets of regular data?

10.4 Using a different random series

Table 8.4.1 listed 12 suitable multipliers, judged on the criterion of having a sigma-coefficient for a moving period of 1 between -90 percent and -110 percent, ±10 percent off the worst -100 percent relationship, the one without any regularity at all. From this list, four other multipliers (28, 11, 7, 2) were tested against the trend model (origin 450, slope 10). The raw data are listed in Appendix 8, and the results are shown in Table 10.4.1.

Table 10.4.1: The results of fitting the trend against varied degrees of randomness

Group (Start)	Trend (450 + 10 × time)							
	Mult 28 (-105%)		Mult 11 (-95%)		Mult 7 (-77%)		Mult 2 (-4%)	
	Sigma	%	Sigma	%	Sigma	%	Sigma	%
All	**30.1006**	100	**28.9011**	100	**28.9377**	100	**28.2459**	100
Small	15 data							
(1)	82.2538	273	30.6692	106	31.8408	110	35.6320	126
(21)	98.8587	328	70.2473	243	39.1844	135	39.7000	149
(41)	49.8100	166	28.5005	99	129.5482	448	98.6560	347
Average	76.9742	256	43.1390	149	66.8578	231	57.9960	205
Medium	30 data							
(1)	31.9864	106	47.1790	163	40.2204	139	48.0848	169
(21)	32.7391	109	38.1445	132	32.4076	112	38.2836	134
(41)	50.2408	167	38.3821	133	30.4076	106	46.5028	163
Average	38.3221	127	41.2352	143	34.3452	119	44.2904	157
Large	60 data							
(1)	30.2110	100	42.8331	148	30.7400	106	59.6575	210
(21)	31.8031	106	28.5041	99	35.5076	123	28.7255	101
(41)	31.0303	103	34.8186	120	29.8163	103	42.0328	148
Average	31.0181	103	35.3853	122	32.0213	111	43.4719	154

None of these series supports the suggestion that 30 observations (shaded in Table 10.4.1) suffice to retrieve the true trend coefficients. In fact, 60 data are barely enough to retrieve the trend when a random factor with multipliers 11 or 2 were used, still producing a sigma as large as 35.38 and 43.47 respectively. This would seem to contradict the previous conclusion. However, one must remember that these series, although having similar standard deviations, shown in bold, are *not* as random as those generated from a multiplier of 18 (their correlations to $M=1$ being shown in parenthesis at the top of each column).

The new series being *not* as random as the old one, paradoxically enough, does corroborate the previous conclusion. To understand this, one must remember that trend data are comprised of two parts: firstly, the true process and, secondly, the random element. Therefore, *the more regular the random element*, the more it *changes* the true process, and *vice versa*. If the true process was changed, then the known coefficients used to measure the success of the model, are no longer the *true* coefficients. Testing for the retrieval of these coefficients does not give an accurate reflection of the number of data required for the retrieval of the trend.

All the random data used have an average of 50.5. This means that although the distribution of numbers may vary from series to series, there are as many numbers above this level as below it, when all data are considered. Therefore, after 100 observations, the effect of changing the true process is neutralised because this is the end of the generated random series. The fit when using 100 observations is close to the true coefficients, not because of the quantity of data, but because the random element has been neutralised over one full cycle.

10.5 Summary

The tests conducted confirm the basic conclusion, namely, that the number of data required to recover a model is exponentially related to the power of that model. This rule applies as long as the data contain a truly random variation. However, if the so-called random element is not random, but contains pockets of regularity, more data may be required. On the other hand, given a certain degree of randomness, a change in the magnitude of the random element does not affect the number of data required to obtain a good fit.

CHAPTER 11

The attempted application

11.1 Introduction

From the previous chapters, two important criteria have been established as prerequisites for accurate forecasting:

- accurately recorded data with individual time intervals;
- sufficient quantity of data to justify a particular model.

Neither of these prerequisites is met by the data storage practices of the client. This does not bode well for a successful attempt at forecasting future sales. Nevertheless, some kind of justifiable future sales level *must* be forecast somehow, to control inventory and to set production levels. Imperfect as they are, all that is available are the data as they stand, and one has to make the best of these in the light of established forecasting techniques and the knowledge gained in this thesis.

What remains is to test the given data by fitting them with their corrected time unit, and restricting the model used to that which the number of observations can reasonably justify.

11.2 The highest model

The periods for which data were provided were two years for Detergents, that is, 24 *monthly* observations, and five years for Personal Products, 51 observations (36 *monthly, 15 other*). Ideally, it would seem logical to use half the available data in fitting, and the other half to test this fit. This also allows enough data for the adequate testing of revision and updating. One would also want the opportunity to search for seasonality. Given that the data are monthly totals, one year represents a complete cycle. Because one should have three times the period length to determine the periodicity, this would require three *years* of data to include the maximum periodicity of twelve. In general, the data available fall short of these basic requirements, but it remains to be seen what can be done.

Expressing the formula for the number of observations required for a certain model in terms of *t*, computes the highest model that should be attempted relative to the given number of observations (formula established in section 9.3).

$$N = 15 \times 2^t$$

$$N/15 = 2^t$$

$$\log (N/15) = t \times \mathrm{Log}\ 2$$

$$\log N - \mathrm{Log}\ 15 = t \times \log 2$$

$$t = (\log N - \log 15) / \log 2$$

With a two year period for Detergents, there are 24 observations. The highest model that should *safely* be attempted is for a time exponent of *(log 24 - log 15) / log 2* = 0.6781. Similarly, for Personal Products, where there are 51 observations, the highest model that should be tested is where time has an exponent *of (log 51 - log 15) / log 2* = 1.7655. Only the whole numbers are significant in these calculations, because one is trying to estimate the *order* of a polynomial. Therefore, the highest model for Personal Products is the trend, and for Detergents, the average.

11.3 Fitting methodology

The number of data available limits the model that should be attempted. Within this limit though, the *amount of improvement* of fit, or the reduction in forecast errors (section 3.2), can justify using a higher model rather than a lower one. This improvement determines whether the model is to include more polynomial or wave terms, or even whether faster revision is better. The standard deviation is a good indicator of how much data can change, and comparing the residual variances of a lower model and a higher one, as a percentage improvement, shows whether the higher model is better or worse.

Computing the residual variance of the raw data with the true average for the observations, sets a base level, the fit of the simplest model with *all* the available data. If a more complex model is to be considered, it should show a positive improvement over using this average. The minimum percentage for upgrading a simple model to a higher one, needs to be considered.

Small improvements in residual variance now, do not necessarily mean better forecasting later. They are computed when fitting and testing with *past* data. A small improvement is not worth capturing because of the associated danger of over-fitting. For the purposes of this study, a 5 percent improvement will be considered significant, ignoring the models that only produce a few percent improvement.

In addition to a significant improvement in fitting, the best model is the one that requires the least revision (section 3.1). A large smoothing factor (alpha) is an indicator that a higher order model should be attempted. If a higher order model fails to *really* improve matters (by at least 5 percent), then one may have to be content with rapid updating because this is less responsive than using a higher order polynomial model. As far as waves are concerned, those with relatively large amplitudes are more important because they describe *real* oscillations rather than small fluctuations. The minimum amplitude considered significant should be the same as for a higher polynomial model, because it also involves additional terms in the model. The amplitude should, therefore, be at least 5 percent of the average for the data.

Fitting the data starts with the simple average, and then proceeds to test increasingly larger smoothing constants. Revision should also improve the forecast by at least 5 percent to be considered significant. Lastly, the detected waves can be tested. Thereafter, the next order of polynomial is added, and the fitting process starts over for this new polynomial. The number of coefficients must not result in over-fitting, and there must be at least a 5 percent improvement to justify extending the model.

11.4 The results

All of the data received from the client are listed in Appendix 9. Product names have been disguised at their request. The *D* or *PP* in each column indicate whether the product is a Detergent or Personal Product. Each separate product has a different letter, and each derivative of the same product, still a separate stock keeping unit, has a different number.

There are 51 observations available for Personal Products, 30 of which will be used for fitting, and the remaining 21 for testing that fit. The success of a model is judged on its ability to reduce the residual variance (mean squared deviation), as compared to the variance for the raw data. For example, the raw data for product PPA1 have a variance of 47435140.5037. However, using a smoothing weight of alpha equal to 11.7647, the variance of the errors is reduced to 10085847.416. To compute an *improvement*, or the reduction in variance, one subtracts this residual variance from the variance of the raw data. As a *percentage* improvement, one divides by the variance of the raw data and multiplies by 100. In other words:

$$(\text{Variance of raw data - Variance of residuals}) \div \text{Variance of raw data} \times 100$$

$$= (47435140.5037 - 10085847.4164) \div 47435140.5037 \times 100 = 78.74\%$$

Those improvements which were negative, meant that that model did not produce more accurate forecasts than the average for all the data, and so had a higher residual variance than the variance of the raw data. In this case, the only model that should be used is the simple average, because any response to the data makes the forecast worse. Indeed, using all the data is better than using only a section of them.

The results for fitting the Personal Products data are shown in Table 11.4.1. The blanks, for example, the periods identified or the size of alpha, meant that the best model did not require any waves or revision.

Table 11.4.1: Personal Product results

Product	Polynomials	Periods	Alpha (%)	Improvement (%)
PPA1	1		11.7647	79
PPA2	1		11.7647	65
PPA3	1		22.2222	68
PPA4	1		11.7647	58
PPB1	2			65
PPB2	2			53
PPB3	2	4	11.7647	**-7**
PPC1	2	3, 2		**-8**
PPC2	1		66.6667	50
PPD1	1	5	11.7647	66

Product group PPA shows a real improvement on the simple average when it is revised, although only by a small alpha. The use of this model is further substantiated by the solidarity between its different derivatives (PPA2 and PPA4), although PPA3 required more rapid revision. Therefore, the simple average with some slight revision in terms of the errors made, is likely to provide a fair estimate of the future values for this product.

However, product group PPB seems to support a trend, although not as convincingly as group PPA supports the average. The exception is PPB3, with -7 percent improvement (shown in bold), which does not support this model, and would be more suited to the simple average for all the data. This can be understood if one knows that this is in fact a bulk product, which is obviously not bought with the regularity of the smaller weighing PPB items.

PPC1 shows no improvement on the average for all the data (shown in bold), whereas PPC2 requires very rapid updating of the average for a significant 50 percent improvement. Although there are 30 data used in the fitting period, and very rapid revision is required, the trend, even with rapid updating itself, does not produce an improvement of this magnitude. This suggests that this rapid revision may only be improving the fit over this particular period, and that the simple average should be used to forecast for both product PPC1 and PPC2.

PPD1 is the only Personal Product tested that showed an improvement from fitting a wave, that is, a 5-point wave. The 30 data used in fitting are sufficient to determine this periodicity, however, bearing in mind that a full cycle for the data is 12 observations or a year, one should be using 36 data to determine the optimal periodicity. Indeed, using 36 of the available 51 data, eliminates the use of a 5-point wave. Product PPD1 is thus best forecasted by the average with slight updating.

Table 11.4.2 shows the results when fitting the first 15 of the 24 observations for the Detergents. These results were then tested on the following 9 observations.

Table 11.4.2: Detergent results

Product	Polynomials	Periods	Alpha (%)	Improvement (%)
DA1	1			-22
DA2	1	2		-46
DB1	1	0		-85
DC1	1	3	40.0000	**19**
DC2	1	3	11.7647	**10**
DC3	1	3	66.6667	**20**
DC4	1	4		**26**
DD1	1	5		-24
DD2	1	3		-26
DE1	1	3, 2	66.6667	-28
DF1	1	3, 2		-50
DF2	1	0		-87
DG1	1	0	11.7647	**54**

The results for the Detergents show a picture very different from the Personal Products. Nearly all the forecasts are worse-off when a higher model than the average was tested. Those with improvements are shown in bold in Table 11.4.2. Product DC1 shows some improvement, but not much consistency as concerns the size of the optimal alpha. Furthermore, the periodicity should only be determined when three full years of data are available. The greatest improvement is shown with product DG1, with the forecast benefiting greatly from some slight revision (alpha = 11.7647).

The poor results with the Detergents suggest that these data have no real pattern, partly because they do not, and partly because there are only 24 observations available. Only the average should be used to make forecasts because there are insufficient data at this point to justify using another model. As many data as are available must be included in the average, so that it is not prejudiced by recent fluctuations.

11.5 Summary

Although there are more data available for Personal Products than Detergents (enough to support a trend), the simple or revised average was the best in all cases except with product group PPB. This shows that generally, there is no real pattern to the data, and that one should not waste time and money trying to discover one.

From the results of the Detergents, it is clear that there is insufficient data in hand to determine whether any true process does exist. Those products which benefit from revision may reveal trends or seasons if more data were available, but in the absence of such data, one must make use of the average and revise it where necessary.

The somewhat disappointing results of investigating the client's data do have one positive conclusion: *there is no point in pursuing sophisticated forecasting procedures until the data available are of sufficient quantity and quality*. The average (simple or smoothed) can be performed by a simple computer program, and expenditure on advanced hardware and software can be avoided.

CHAPTER 2

Algebraic time series models

2.1 Introduction

Forecasting involves recognising the characteristics of the process which resulted in a given set of data. The *true* process characteristics are those that prevailed in the past, and are likely to continue in the near future. The characteristics of data can be grouped two into categories.

Firstly, those that have a structure, pattern or repetitive tendencies. In this case, discovering these characteristics will provide objective parameters for forecasting future values.
Secondly, data that have no structure, where there is no recognisable relationship between successive values: these are called *random* data.

Random data are characterised by a lack of structure and, therefore, do not allow the forecasting of specific future values. It is, however, possible to still forecast a range of values, limits which the data have not exceeded in the past, and are unlikely to exceed in the future. Knowledge of these limits is more advantageous than not knowing them, but, more importantly, recognising that the data are random means that time is not wasted searching for structures that are not there.

2.2 The single most important characteristic

When summarising a given set of data, one is looking for characteristics which are more informative than simply listing the data. Initially, one may only be interested in the *single* most dominant characteristic, for example, whether the data are "high" or "low". So which single number provides a good idea of how the data are behaving as a whole?

One could use *any* one datum from the current series. Picking the first or last number may seem arbitrary, but it is intuitively better than doing nothing at all to plan for the future. However, it is not obvious why such a choice would be better than any other arbitrary selection, such as one from the middle of the set, or the most frequent datum.

One objective method, based on simple arithmetic and the data in hand, would be to calculate the discrepancy between the estimate and each datum, and then the total discrepancy. The lower the total discrepancy, the more representative the estimate is. This procedure for testing an estimate against a set of data can be thought of as *fitting*, just like trying on new clothes before deciding which looks best. The discrepancies appear as errors, the *mistake* made when fitting the estimate to the data. This procedure will initially be a manual trial and error process, to be expanded into an objective, mechanised routine.

Fitting can be illustrated with a series of simple data, for example, the series 10 10 10 15 10. As the first estimate or *fit*, one might try the mode, or most frequent datum which is obviously 10, but one could also have chosen the first or last number as it happens. Table 2.2.1 shows how the fit is tested and how the errors are computed.

Table 2.2.1: Fitting the mode (10) to the data

Time (t)	Data (D)	Fit (F)	Error = Data - Fit (Err)	Error squared (Err^2)
1	10	10	10 - 10 = 0	0
2	10	10	10 - 10 = 0	0
3	10	10	10 - 10 = 0	0
4	15	10	15 - 10 = 5	25
5	10	10	10 - 10 = 0	0
Sum	55		5 |5|	25

The abbreviations in parenthesis, in Table 2.2.1, such as "t" for "time" and "D" for "Data", will be used in later tables instead of the headings. As previously stated, the *error* is the difference between the actual data and the fit. Squaring these *simple* errors, shown in the last column, increases large errors relatively more than smaller ones, and also eliminates any negatives. This is computed to lead subsequently to a comparison between the sum of the simple errors and those squared. The last row in Table 2.2.1 is the sum of each column. There are two sums for the error column. Firstly, the sum of the simple errors, and secondly, the sum of the *absolute* errors, placed between bars. The absolute error is the deviation regardless of whether it is positive or negative.

In the case of the mode, the sums of absolute and simple errors were identical. The fit was perfect on every day, except day 4, where there was a large error of 5. If a somewhat larger fit were tested than the previous one, say 11 instead of ten, how would this change the resulting errors?

Table 2.2.2: Fitting 11 to the data

t	D	F	Err	Err^2
1	10	11	-1	1
2	10	11	-1	1
3	10	11	-1	1
4	15	11	4	16
5	10	11	-1	1
	55		0 |8|	20

In Table 2.2.2, the sum of the simple errors has been reduced to 0. This is significant because it shows that the fit passes through the centre of the data, there being as many positive errors as there are negative ones. In other words, the fit over-estimated the data as often as it under-estimated them. This was not the case with the previous fit, the mode in Table 2.2.1, with a simple error of 5.

On the other hand, if the sum of the simple errors is zero, this does not mean that there were no errors. This suggests that the sum of the simple errors is not in itself, or on its own, a valid criterion for measuring the goodness of a fit. A second criterion is needed, either the sum of the absolute errors, or the squared ones. Considering the results of an even larger fit, say 12 instead of 11, will help clarify how these criteria change as a function of the fit.

Table 2.2.3: Fitting 12 to the data

t	D	F	Err	Err^2
1	10	12	-2	4
2	10	12	-2	4
3	10	12	-2	4
4	15	12	3	9
5	10	12	-2	4
	55		-5 $\lvert 11 \rvert$	25

The fit of 12 in Table 2.2.3 results in a smaller simple error of 3 on day 4, *versus* 4 and 5 previously, but larger errors on the other days of -2 *versus* -1 and 0. A summary of the sum of various errors from fitting 10, 11 and 12 is represented in Table 2.2.4.

Table 2.2.4: Summary of errors when fitting a constant

F	$\sum$Abs Err	$\sum$Err	$\sum Err^2$
10	5	5	25
11	8	0	20
12	11	-5	25

Table 2.2.4 shows that the fit of 11 has the lowest sum of the simple errors ($\sum$Err), and also the lowest squared errors ($\sum Err^2$), while the fit of 10 has the lowest sum of the absolute errors ($\sum$Abs Err). Deciding which fit is the *best* therefore depends on the criterion considered most important.

With the simple error, a less than perfect fit (for example 11), can still have a total error of zero. This is because the sum of simple errors ignores the nature of negative and positive errors. In the context of manufacturing, over supply and under supply of a product do not have the same effect on costs, yet both are serious and it is the aim of the forecaster to eliminate such wastage. At least the absolute error considers this, but unfortunately it does not discriminate between large and small errors. As a rule, large errors are much more costly than small errors, whether they be positive or negative. Large errors should, therefore, be given higher weighting. Indeed, such weighting can be achieved by multiplying an error by itself, *squaring* it.

The sum of the squared errors has two important properties. Firstly, by squaring each error, the negatives become positive. Secondly, squaring effectively assigns more weight to large errors than to small ones, with equally negative and positive errors being given the same weight. In principle, all errors are given some weight. Thus, the *least squares* criterion, where the best forecast is judged on its ability to minimise the sum of the squared errors, seems intrinsically meaningful. Moreover, choosing the squared error criterion over the absolute one, leads to a *method* of computing the best fit, the one minimising the squared errors. With the absolute error, the only method is trial and error.

Using tables to establish the sum of the squared errors for a series of data is convincing, but impractical, even for the simple data used thus far. However, studying the computational processes behind the sums in the previous tables leads to the *least squares method* just mentioned.

Table 2.2.5: Developing a formula for the sum of the squared errors

t	D_t	F	Err^2
t_1	D_1	F	$(D_1 - F)^2$
t_2	D_2	F	$(D_2 - F)^2$
t_3	D_3	F	$(D_3 - F)^2$
t_4	D_4	F	$(D_4 - F)^2$
t_5	D_5	F	$(D_5 - F)^2$

From the shaded column in Table 2.2.5, a formula for the sum of the squared errors can be derived as $\sum Err^2 = \sum (D_t - F)^2$. This can be expanded:

$$\sum Err^2 = \sum (D_t - F)^2 = \sum (D_t - F)(D_t - F) = \sum (D_t^2 - 2D_t \times F + F^2)$$

This can be simplified by using the principle that the sum of a series, each element of which is multiplied by a constant, is equivalent to the sum of that series, multiplied by the constant. *N* is the number of data in the series.

Principle: $\sum(constant \times data) = constant \times \sum data$, the fit being the constant

Therefore: $\sum D_t^2 - 2F \times \sum D_t + N \times F^2$

There is thus a formula for computing the sum of the squared errors, when a constant level (single estimate) is fitted to the data. This can be tested against the results achieved in Tables 2.2.1, 2.2.2 and 2.2.3. In order to compute the sum of the squared errors with the new formula, values for the sum of the data, and the sum of the data squared are required. Table 2.2.1 is repeated in Table 2.2.6, with an extra column added for data squared, and the sums for the data and data squared.

Table 2.2.6: Testing the formula with the mode (10)

t	D	D^2	F	Err	Err^2
1	10	100	10	0	0
2	10	100	10	0	0
3	10	100	10	0	0
4	15	225	10	5	25
5	10	100	10	0	0
	55	625		0 |5|	25

$$\sum Err^2 = \sum D_t^2 - 2F \times \sum D_t + N \times F^2 = 625 - 2(10) \times 55 + 5 \times (10)^2$$

$$= 625 - 1100 + 500 = 1125 - 1100 = \mathbf{25}$$

The answer of 25 is the same as in Table 2.2.1. The other fits can also be tested.

Fit of 11: $\sum Err^2 = 625 - 2(11) \times 55 + 5 \times 121 = \mathbf{20}$

Fit of 12: $\sum Err^2 = 625 - 2(12) \times 55 + 5 \times 144 = \mathbf{25}$

These results show that the formula for the sum of the squared errors has been derived correctly because it gives the same answers as the tables did. Consideration of other fits that both under estimate and over estimate the data, yields the results in Table 2.2.7.

Table 2.2.7: $\sum Err^2$ when fitting a level from 9 to 15

F	9	10	11	12	13	14	15
$\sum Err^2$	40	25	20	25	40	65	100

It can be seen from Table 2.2.7 that the lowest sum of the squared errors is where a fit of 11 is used. Based on this criterion, 11 is the *best* fit for the data. The values for the sum of data and data squared, computed by hand in Table 2.2.6, can easily be computed routinely from any set of data. Although one can now compute the sum of the squared errors without using a table, determining the *best* level fit from those tested was done visually in Table 2.2.7.

A formula minimising the sum of the squared errors computes F^*, the best level or *horizontal fit* for the data. This can be done by taking a derivative with respect to F, and then solving for F^*. The formula for a derivative is: $X^n = n \times X^{n-1}$

Looking at the formula $\sum D_t^2 - 2F \times \sum D_t + N \times F^2$, the first term ($\sum D_t^2$) does not affect the fit (F). It can therefore be excluded when taking the derivative (d) with respect to F.

Therefore:

$$\frac{d \sum Err^2}{d F} = -2\sum D + 2N \times F$$

To minimise the sum of the squared errors, the derivative is set to zero and solved for F^*:

$$-2\sum D_t + 2N \times F = 0, \text{ minimising condition}$$

$$-\sum D_t + N \times F = 0, \text{ simplified by dividing by 2}$$

$$N \times F = \sum D_t, \text{ unsolved form}$$

$$F^* = \sum D_t/N, \text{ solved}$$

Substituting the original data to test the formula for minimising the sum of the squared errors gives: $F^* = 55/5 = 11$. Indeed, 11 is the *best* horizontal fit for the data 10 10 10 15 10.

2.3 Constant algebraic model

The formula for F^* is the sum of all the data in the series, divided by the number in the series. In other words, the simple average or *mean* is the best single estimate of the data. The average, also termed the *constant algebraic model*, principally computes a level midway between the highest and lowest data in the series. Therefore, the sum of the simple errors will always equal zero, but more importantly, the sum of the squared errors will be minimised while giving more weight to large errors. This model provides a methodology for establishing the *best horizontal fit* objectively and repeatedly.

It also provides a basis for fitting higher order models.

2.4 Linear algebraic model

The data used to forecast with the *constant algebraic model*, 10 10 10 15 10, did not change from one to another with any regularity. The data 10 20 30 50 40 can be seen to be increasing in value with time, at least for the first few values. The *rate of change* defines by how much one datum changes from the previous one. There are negative and positive rates of change, depending on whether the data are descending or ascending. The rates of change for the data 10 20 30 50 40 are shown in Table 2.4.1.

Table 2.4.1: The rate of change for the given data

t	D	Rate of change (ROC)
1	10	-
2	20	10
3	30	10
4	50	20
5	40	-10

The rate of change for these data is a constant 10 for $t=2$ and $t=3$. The rate of change changes for $t=4$ and $t=5$. There is no rate of change for $t=1$ because there was no previous datum to compare it to. If a forecast were to be made using the *constant algebraic model*, the best level would be 30, but the sum of the squared errors would be 1000. Large errors result on each day except on day 3, where there is a perfect fit, because the average ensures that the level passes through the middle of the data. However, the best level ignores the rate of change for these data, and so it does not provide a good fit. In the face of changing data, a model is required that will keep up with this change. One could utilise a trend model.

However, a trend has a constant rate of change, because it is a straight line, and the data have a changing rate of change. This raises the issue of which of the rates of changes in Table 2.4.1 to select. One possibility is to use the most common rate of change. A simple method for fitting a trend to data, is to add the selected rate of change (10) to the average for the data (30) as time progresses, and subtract it from the average going back in time from the average. The results are shown in Table 2.4.2, the average having been shaded.

Table 2.4.2: Fitting a trend with a rate of change of 10 to the sample data

t	D	F (ROC=10)	Err	Err^2
1	10	10	0	0
2	20	20	0	0
3	30	30	0	0
4	50	40	10	100
5	40	50	-10	100
	150		0 |20|	200

Simply by fitting a trend with an arbitrary rate of change, instead of the average, the sum of the squared errors has been reduced from 1000 to just 200 in Table 2.4.2. This is because the trend *captures* the change that occurs between the first three data perfectly, and some of the change that occurs in the data later.

However, this may not be the *best* trend to fit, because the rate of change of the data is not constant. The average rate of change, is the simple sum of the changes (30, in Table 2.4.1) divided by the number of changes (4), which is equal to 7.5. Therefore, on the average the data change from one day to another by 7.5. As this change is lower than 10, a smaller rate of change may provide a better fit.

Table 2.4.3: Fitting a trend with a rate of change of 9

t	D	F (ROC=9)	Err	Err^2
1	10	12	-2	4
2	20	21	-1	1
3	30	30	0	0
4	50	39	11	121
5	40	48	-8	64
	150		0 |22|	190

With a rate of change of 9 in Table 2.4.3, the sum of the squared errors has been reduced to 190. Although the rate of change is less than that for the first three data, and small errors result, the sum of the squared errors was reduced from 200 to 190. A trend with an even smaller rate of change may further reduce the sum of the squared errors.

Table 2.4.4: Fitting a trend with a rate of change of 8

t	D	F (ROC=8)	Err	Err^2
1	10	14	-4	16
2	20	22	-2	4
3	30	30	0	0
4	50	38	12	144
5	40	46	-6	36
	150		0 |24|	200

The even lower rate of change of 8 in Table 2.4.4, has pushed the sum of the squared errors once again to 200. The reduced error at $t=5$, because of the lower rate of change, is outweighed by the larger errors incurred on the other days, because of this same lower rate of change. Thus far, the best fit has been for a rate of change of 9.

Using tables to establish the best rate of change, is an inefficient process. To improve this, one needs a formula to compute the sum of the squared errors, as well as the best origin and rate of change, so as to minimise the sum of the squared errors.

From basic mathematics, any trend can be represented by a formula. It includes the origin and the rate of change, or the slope multiplied by time, $Fit(t) = A + B \times t$. The origin is the data value at time zero (t=0). From Table 2.4.3 with a rate of change or slope of 9, the Fit(t=0) or origin is 12 (t=1) minus 9 (slope) which equals 3. Therefore, in the formula for a trend, $A=3$ and $B=9$. Instead of fitting the trend as done in Table 2.4.3, the trend is fitted using this formula in Table 2.4.5.

Table 2.4.5: Fitting the trend where A=3 and B=9

Origin (A)	Slope (B)	t	D	A + Bt	Err^2
3	9	0	-	3	-
3	9	1	10	12	4
3	9	2	20	21	1
3	9	3	30	30	0
3	9	4	50	39	121
3	9	5	40	48	64
			150		190

The tables are identical, including the sum of the squared errors, except for the data values for t=0, the origin. From Table 2.4.5, the formula for the squared error is still $(Data - Fit)^2$ for each point in time, except that the fit is not simply F as before for the constant model. Rather, it is the trend formula, $A + B \times t$.

The formula for the squared error for a trend now reads:

$$Err^2 = [Data - (A + B \times t)]^2 = (D_t - A - B \times t)^2$$

This can be expanded: $(D_t - A - Bt) \times (D_t - A - Bt)$

$$= D_t^2 - AD_t - BtD_t - AD_t + A^2 + ABt - BtD_t + ABt + Bt^2$$

$$= D_t^2 - 2AD_t - 2BtD_t + 2ABt + A^2 + Bt^2$$

This is the formula for the squared error at any one point in time. It can be expanded to include the sum of all the squared errors, for time from 1 to n. Table 2.4.6 applies the principle that the *sum of a series multiplied by a constant is, equal to the sum of that series, multiplied by the constant.* N is the number of data.

Table 2.4.6: The sum of the squared errors over an interval t=1,n

t	D_t^2 - $2AD_t$ - $2BtD_t$ + $2ABt$ + A^2 + Bt^2
1	$D_1^2 - 2AD_1 - 2B(1)D_1 + 2AB(1) + A^2 + B(1)^2$
2	$D_2^2 - 2AD_2 - 2B(2)D_2 + 2AB(2) + A^2 + B(2)^2$
3	$D_3^2 - 2AD_3 - 2B(3)D_3 + 2AB(3) + A^2 + B(3)^2$
4	$D_4^2 - 2AD_4 - 2B(4)D_4 + 2AB(4) + A^2 + B(4)^2$
5	$D_5^2 - 2AD_5 - 2B(5)D_5 + 2AB(5) + A^2 + B(5)^2$
$\sum Err^2 =$	$\sum D_t^2 - 2A\sum D_t - 2B\sum t \times D_t + 2AB\sum t + NA^2 + B\sum t^2$

From Table 2.4.6, a formula can be derived, to define the sum of the squared errors for any origin and slope. Moreover, in keeping with the least squares methodology, one can directly minimise the sum of the squared errors.

Firstly, taking the derivative of this formula with respect to A, then to B, yields the formulae for origin and slope.

$$\sum Err^2 = \sum D^2 - 2A\sum D - 2B\sum tD + 2AB\sum t + N \times A^2 + B\sum t^2$$

$$\frac{d \sum Err^2}{d\,A} = -2\sum D + 2B\sum t + 2AN \qquad \frac{d \sum Err^2}{d\,B} = -2\sum(tD) + 2A\sum t + 2B\sum t^2$$

Secondly, to minimise the sum of the squared errors, the derivatives are set equal to zero and simplified:

$$\frac{d \sum Err^2}{d\,A}: -2\sum D + 2AN + 2B\sum t = 0 \qquad \frac{d \sum Err^2}{d\,B}: -2\sum(tD) + 2A\sum t + 2B\sum t^2 = 0$$

$$2AN + 2B\sum t = 2\sum D \qquad 2A\sum t + 2B\sum t^2 = -2\sum(tD)$$

$$\mathbf{AN + B\sum t = \sum D} \qquad \mathbf{A\sum t + B\sum t^2 = \sum(tD)}$$

By taking the derivatives for A and B from the sum of the squared errors formula in Table 2.4.6, and setting them to zero, the lowest sum of the squared errors is guaranteed. There are now formulae for the *best* origin and slope, A^* and B^*.

As a matter of consistency, one can show that in the event that there is no slope (B=0), then the best fit is A^* (because the data follow no particular trend).

A^* (B=0): $\quad AN + B\sum t = \sum D$

$A \times N + 0\sum t = \sum D \qquad A = \sum D/N$, the mean, as before.

The trend model is consistent with the *constant algebraic model*, because in the absence of any slope, the formulae are the same. The trend simply expands the average model, by including a slope in the fit.

In order to determine the best trend fit with the data 10 20 30 50 40, the two equations for A^* and B^* need to be solved simultaneously. The equations are:

$$A \times N + B \times \sum t = \sum D \text{ and } A \times \sum t + B \times \sum t^2 = \sum (t \times D)$$

Their solution requires values for $\sum t$, $\sum D$ and $\sum t^2$, that is, the sum of time, of the data, and of time squared. These are computed in Table 2.4.7. N is equal to 5.

Table 2.4.7: Data to be used for simultaneous equation solution

t	D	t × D	t^2
1	10	10	1
2	20	40	4
3	30	90	9
4	50	200	16
5	40	200	25
15	150	540	55

Substituting from Table 2.4.7: $A \times 5 + B \times 15 = 150$

$A \times 15 + B \times 55 = 540$

The substitution method for solving simultaneous equations results in the solutions of $A^*=3$ and $B^*=9$, the best origin and slope, as shown in Appendix 1. This solution is consistent with the trend visually estimated from Table 2.4.3. It computes the *best* trend fit by minimising the sum of the squared errors for all possible slopes, not just for those selected arbitrarily. These formulae also greatly improve computational efficiency.

The *linear algebraic model* is a model that fits the *best* trend, according to the least squares method, for any series of data. This is suited to data having increasing or decreasing tendencies, allowing the forecast to *catch* on to this trend. When there is no slope, the best "trend" model reverts back to the *constant algebraic model*. This is obvious in cases of random data, where individual data should not have any relationship to those before and after them. In other words, random data should have a slope of zero. If they do not, they are either intrinsically not random, or else the sample is too small, one having fallen upon a pocket of non-randomness or regularity, in a series that might well turn out to be random in the long run.

2.5 Polynomial algebraic model

The data 4 6 10 16 24 34 have a rate of change which increases with time, that is, the rate of change itself changes. One can see this when studying the rate of change of the rate of change, as illustrated in Table 2.5.1. With the data increasing as rapidly as they are, both the average and the trend models would produce large errors. This is because no matter what level or slope was used, the data are increasing at an increasing rate and a simple forecast will always fall short.

Table 2.5.1: The changing rate of change

t	D	ROC	ROC of ROC
0	4	-	-
1	6	2	-
2	10	4	2
3	16	6	2
4	24	8	2
5	34	10	2

The changing rate of change shown in Table 2.5.1 can be correctly estimated by deriving the formula from first principles, as was done before for the linear model. However, one can also study the formulae already derived, with a view to finding a system in the progression of terms.

$$A \times N + B \times \sum t = \sum D \qquad \Longrightarrow \qquad A \times \sum t^0 + B \times \sum t^1 = \sum (t^0 \times D)$$

$$A \times \sum t + B \times \sum t^2 = \sum (t \times D) \qquad \Longrightarrow \qquad A \times \sum t^1 + B \times \sum t^2 = \sum (t^1 \times D)$$

Firstly, time can be consistently written with its proper exponent. Time t can be written as time with an exponent of one, t^1. Similarly, the sum of the time, $\sum t$, can be written as $\sum t^1$. Time with an exponent of zero, t^0, is equal to one, therefore, D multiplied by t^0 is the same as D. Furthermore, the sum of time with exponent zero, $\sum t^0$, is $1 + 1 + 1 \ldots N$ times, so it yields the number of data in the series, N.

Expanding the power of t seems to complicate otherwise simple equations, but in fact brings out the regularity of the terms. Now that all powers appear explicitly, it is possible to derive by inspection what the next higher coefficient should look like:

$$\begin{array}{llll} A \sum t^0 + & B \sum t^1 + & C \sum t^2 & = \sum (t^0 \times D) \\ A \sum t^1 + & B \sum t^2 + & C \sum t^3 & = \sum (t^1 \times D) \\ A \sum t^2 + & B \sum t^3 + & C \sum t^4 & = \sum (t^2 \times D) \end{array}$$

The visual progression adds a new column for C, and a new row starting with time squared. On the other hand, as a matter of consistency, one reverts to the simple mean if both B and C are zero, that is, $A \sum t^0 = \sum (t^0 \times D)$, therefore $A = \sum (t^0 \times D) / \sum t^0$. Similarly, if C is zero, one reverts to the trend model. But to determine whether the best rate of change of the rate of change should be zero or not, all three equations must be solved. The equations are ordered with respect to the exponent of time in the first column. For example, the trend is a first order polynomial, and a trinomial is a second order polynomial. The average is the special case of a zero order polynomial.

It can thus be seen from the above equations, that the *constant algebraic model* can be extended to the *linear algebraic model*, which can be extended to the *curved algebraic model*, which is extendable to a polynomial of any order. In general, "...one can use any degree polynomial that is required to represent the process by adding terms in t^2, t^3, and so on, up to t^N..." (Brown, 1963: 62).

It should be noted that the final rate of change in Table 2.5.1, the *ROC of ROC*, is a constant of 2. As mentioned in section 1.2, the actual rate of change could reflect a percentage growth. This process requires an exponential model to fit it (Brown, 1963: 64). It is, however, more likely to be relevant to biological processes rather than to manufacturing.

2.6 Summary

The *constant*, *linear* and *polynomial algebraic models* have been shown to be consistent with one another, and to be easily expanded. One can even achieve a perfect fit with any data by having as many terms as there are data, which would be a purely mechanical process. Having such a high degree model, although fitting the data perfectly, does not assist in forecasting future values, but rather inhibits it because the model continues on its final rate of change. This is called *over-fitting*, where a model has too many terms. It may obtain a perfect fit without reflecting the underlying processes currently happening in the market.

Over-fitting will be dealt with in the next chapter, by revising a model as new data are received, rather than with simply extending the model to fit them.

CHAPTER 3

Revision in the light of new data

3.1 Introduction

No matter how simple or complex a given model is, it will continue on its current rate of change as time goes by. For example, a trend eventually increases to infinity or decreases to zero and beyond, but it is inconceivable that real data will ever continue to behave in a similar fashion. Increasing the power of the model can lead to over-fitting, the model only fitting the data that have gone before, minimising the errors between the model and the data, but without revealing any true process. In other words, for a model to be *good*, it is not enough that it simply fits the past. It must also continue to fit, without major change, new data as they become available. Indeed, the *best* model is the one that needs the least revision, not having to be changed with every new datum.

So the main idea of *revising* is to adapt a model in terms of *new* data as they occur. This also implies that, as time goes by, older data become more distant, and therefore, less important to what is happening today. Keeping old data also requires storage space, and entails the risk of data being lost, even with good record keeping, through a computer or administration catastrophe. To keep the forecast up to date and accurate, it must be revised to a certain extent in terms of the latest data, yet include enough of the old data to establish a good quality model.

3.2 The moving average

To begin with, consider the simplest possible model, the average, and that the most recent three data are considered significant for a forecast. Only the three latest data would then be summed and divided by 3, meaning that a *moving period* (M) of 3 had been selected.

The data 3 9 5 7 9 2 7 8 8 2 are random numbers taken from Trueman (1981: Appendix E), and can be used to illustrate this. The average for these data, with a moving period of 3 (M=3), can be seen as follows:

Moving average for day 4 =	3 +	9 +	5			= 17 ÷ 3 = 5.7
Moving average for day 5 =		9 +	5 +	7		= 21 ÷ 3 = 7
Moving average for day 6 =			5 +	7 +	9	= 21 ÷ 3 = 7
Moving average for day 7 =				7 +	9 + 2	= 18 ÷ 3 = 6

.
.

The first forecast using the *moving average* can only be made on day 4, when three data are in hand. On each subsequent day, the three most recent data are summed and divided by 3. The general principle of the moving average is:

$$\text{Moving average for day t} = \frac{D(t\text{-}3) + D(t\text{-}2) + D(t\text{-}1)}{\text{moving period}}$$

The first of the three data in the average is the oldest, from three days ago (t-3). With each new day, the oldest datum is dropped and the newest is added, for example, on day 5, 3 is subtracted and 7 is added. The forecast for tomorrow includes two of the data from the previous day's average, plus the datum from today. Each new colour (for example, green) "moves" from the last datum, to the middle, to the first in the series over the three days.

Alternatively, one can subtract the oldest from the newest datum and divide by the moving period, then add the previous average:

$$\text{Moving average for day t} = \text{previous average} + \frac{\text{newest data - oldest data}}{\text{moving period}}$$

Example: Moving average for day 6 = 7 + (9 - 9) ÷ 3 = **7**

The formula can also be seen as follows, where dividing by *M* is the same as multiplying by *1/M*, and the difference between the oldest and the newest data can be seen as the "change" in the data:

$$\text{Fit (t)} = \text{Average} + 1/M\ (\text{New } D_t - \text{Old } D_{t\text{-}M}) = A + 1/M \times \text{"change"}$$

Example: Moving average for day 7 = 7 + 1/3 (2 - 5) = **6**

In other words, a moving average can be interpreted as follows: the forecast for tomorrow is based firstly, on the average for today, and secondly, on a fraction (1/M) of the newest datum, whilst removing the same fraction of the oldest datum. This in turn is equivalent to including *1/M* of the *change* from the oldest observation to the most recent one.

Thus, as the average *moves* through time, it incorporates some of the *change* that occurred in the data during the moving period. One can again compute the sum of the squared errors as done before when testing a fit by trial and error. The results of forecasting with $M=3$ for the random data can be seen in Table 3.2.1.

Table 3.2.1: Results of the moving average (M=3)

t	D	Average	Err	Err^2
1	3	-	-	-
2	9	-	-	-
3	5	-	-	-
4	7	5.7	1.3	1.7
5	9	7	2	4
6	2	7	-5	25
7	7	6	1	1
8	8	6	2	4
9	8	5.7	2.3	5.3
10	2	7.7	-5.7	32.5
				73.5

The computed average is shaded in Table 3.2.1, after three days have elapsed to accumulate the first average. The choice of $M=3$ was selected simply to illustrate the moving average. A total is given in the last row for the sum of the squared errors.

A shorter moving period, such as M=2 or 1, or a longer one, M=4 or 5, could be tested to determine the effect on the sum of the squared errors. All the sums of the squared errors for moving periods from 1 to 5 are shown in Table 3.2.2. N is the number of data on which the average was tested, that is, the number of forecasts that were made. The more data included in the average, the less are available to test the fit. In the last column, the sum of the squared errors is divided by N.

Table 3.2.2: Comparison of squared errors for M=1 to 5

M	N	$\sum Err^2$	$\sum Err^2 / N$
1	9	171.0	19.0
2	8	96.8	12.1
3	7	73.5	10.5
4	6	64.2	10.7
5	5	50.5	10.1

From Table 3.2.2, $M=5$ has the lowest sum of the squared errors (shaded). According to the least squares methodology, $M^*=5$. From Table 2.2.4, the best indicator of goodness of fit was the sum of the squared errors because it gives more weight to large errors, less weight to small ones, but in principle, some weight to all errors. In Table 3.2.2, dividing this sum by the number of data used to test the average (Mean Squared Error) puts the results in perspective. The errors from $M=3$, $M=4$ and $M=5$ are about equally large, not immediately obvious from the squared error totals. A moving period larger than 5 would mean that more than half the data are being used in the average, so there would not be enough data left to adequately test the forecast.

Looking at the formula for the moving average, $Fit\ (t) = Average + 1/M\ (New\ D_t - Old\ D_{t-M})$, the forecast is based on a percentage (1/M) of the change between the newest and the oldest data. As the length of the moving period is changed, so this percentage changes inversely. For example, a moving period of 4 means that the forecast includes 25 percent (1/4) of the change, whereas, if the moving period is halved to 2, 50 percent of the change (double) is considered. This percentage refers to how *responsive* a forecast is with respect to new data. The more responsive the model is, the greater the effect the newest datum has on the average. The longer the moving period, the less responsive the model becomes, because a smaller percentage of the change is considered.

If $M=1$, the forecast for tomorrow is based 100 percent (1/1) on today's datum. This is a special case of the moving average because the forecast for tomorrow *is* the datum from today, as shown in Table 3.2.3.

Table 3.2.3: Moving average (Avg) of M=1

t	D	Avg	Err	Err^2
1	3	-	-	-
2	9	3	6	36
3	5	9	-4	16
4	7	5	2	4
5	9	7	2	4
6	2	9	-7	49
7	7	2	5	25
8	8	7	1	1
9	8	8	0	0
10	2	8	-6	36
				171

Mean Squared Error = 171 ÷ 9 = 19

Table 3.2.3 has the largest mean of the squared errors for all the moving periods tested. This is not surprising because responding to random data only leads one off-course, because they are random. The most responsive forecast should, therefore, be the worst. In this example, it is also to be expected that the longest moving period would be the best because it is least responsive to random data.

As before, a methodology is required that will determine M^* without computing the sum of the squared errors for each moving period. For this purpose, three other statistical indicators of error will be considered which are computed from errors relevant to the number of data available to test the average. These are:

Mean Absolute Deviation (MAD):	*This is the average absolute error, which is the sum of the absolute errors divided by the number of forecasts made.*
Variance (MSD):	*Also referred to as the Mean Squared Deviation (previously the Mean Squared Error), this is the average of the squared errors, or the sum of the squared errors divided by the number of forecasts made.*
Standard Deviation (SD or σ):	*The square root of the Variance (MSD). It is an indicator of how much individual data vary from the mean.*

These definitions can be applied to the example for M=3:

MAD = 19.3 ÷ 7 = 2.76 Variance (MSD) = 73.47 ÷ 7 = 10.5 SD = √MSD = 3.24

The MAD, MSD and SD are shown for all the moving periods in Table 3.2.4.

Table 3.2.4: Comparison of MAD, MSD and SD for M=1 to 5

M	MAD	MSD	SD (σ)
1	3.67	19.00	4.36
2	2.69	12.09	3.48
3	2.76	10.50	3.24
4	2.88	10.70	3.27
5	2.68	10.10	3.18

The moving period of 5, shaded in Table 3.2.4, is best on all counts. According to the *Least Squares Methodology*, minimising the sum of the squared errors is a criterion aimed at reducing larger errors by weighting them more. Squaring makes the sum of the errors too large, but by taking the square root of the *Variance (MSD)*, the *Standard Deviation* is brought back to the same order of magnitude as the *Mean Absolute Deviation*, in spite of having weighted larger errors to a greater degree. The standard deviation (sigma) is, therefore, a valid criterion for judging the accuracy of a particular forecast, M^* being determined by the lowest sigma (σ).

3.3 Consistency of the moving average

There is an internal inconsistency with the moving average. In the case of the random data above, the moving average responded as expected, a longer moving period being best for such data. Similarly, a shorter period would prove to be best if the average was moved through data with a trend or other pattern, and it would be comparatively simple to demonstrate this.

However, before engaging in such a project, one should consider whether this is in fact worth doing. Unfortunately, the moving average is not intrinsically suited to keeping up with the unfolding process, for two reasons. Firstly, recent data are *included* and receive some weight, but older data are *excluded* and receive no weight from the average. Secondly, data within the actual moving period are of *different* ages but they receive *equal* weight.

The older data within the moving period are still treated with the same importance as the new data, because all data are divided equally in an average. Each datum in the moving period is given *equal weight*, and those outside of the moving period are given no weight at all. The graphical display of the weight assignment (M=3) by the moving average can be seen in Figure 3.3.1.

Figure 3.3.1: Weight assignment with M=3

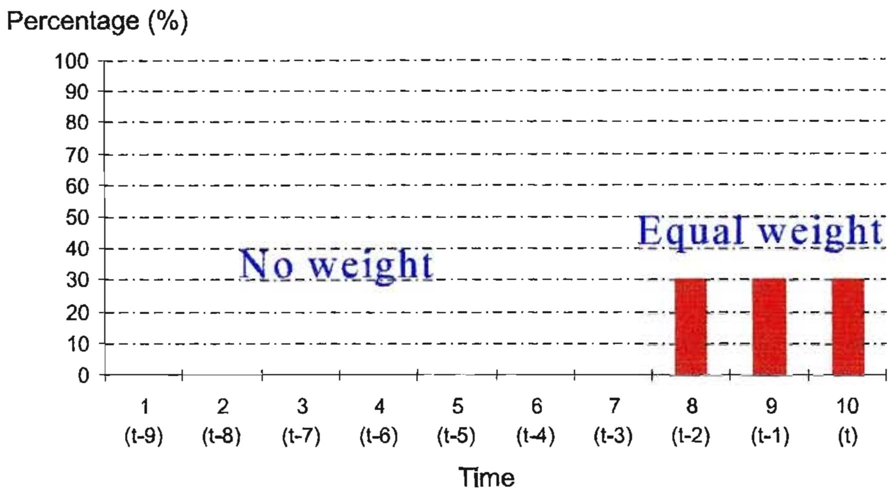


Figure 3.3.1 shows that 33 percent weight is given to each datum in a moving period of 3, and no weight is given to data older than those involved in moving the period. Moving assigns weight to data in the moving period uniformly, and abruptly fails to consider data excluded from this period. It would seem more reasonable to give more weight to recent data, less weight to distant data but, in principle, some weight to all data. Ideally, weight would be assigned according to the age of the data, with older data being of less importance.

To develop a method leading to proportional weights, Brown suggests: "...suppose...there were some catastrophe in the data-processing centre, which destroyed all historical information..." (1963: 100). How would one replace the information lost in such a scenario? There are three plausible answers:

- omit the past data from forecasts;
- replace lost data with a subjective estimate;
- substitute the current average for the lost data.

Omitting data seems arbitrary, while using subjective methods defeats the idea of forecasting, that is, objectivity. Or rather, if one were to replace the missing data subjectively, would one not use the last available average? The formula for the forecast for tomorrow would change such that the *'Old'* section of the *'New - Old'* becomes *'Average'*, the old data having been lost.

Moving:	A = A + 1/M (New - Old)	CHANGE from oldest to newest
After loss:	A = A + 1/M (New - Average)	ERROR between newest and previous average

Now, with each new day, the forecast for tomorrow is the forecast for today (as before), plus a fraction of the difference between today's forecast and today's datum. It is no longer the *change* in the data from oldest to newest that is considered, but today's *error*.

Not only is the new formula meaningful in itself, it is also simpler to compute. All that is required for tomorrow's forecast is today's datum and today's forecast. This not only provides an objective assessment of the lost data, but makes the equation much simpler than before: none of the many past data need be stored at all but only their average, a single value. There is now a new formula in its own right.

In view of the inconsistencies with moving, it is important to investigate how this new formula assigns weight to past data. For the purpose of studying this, in the formula $A = A + 1/M\,(D - A)$, the weight $1/M$ will be replaced by w.

Therefore: $$A = A + w(D - A) = A + wD - wA = wD + A(1-w)$$

Today's data is weighted by w, while the average is weighted by the complement of w, $1\text{-}w$. In order to simplify, the weight (w) for today's data is commonly called alpha (α), and its complement, beta (β). Therefore $A = \alpha D_t + \beta A$. Figure 3.3.2 shows how this formula assigns weight to past data as they become older and older, by investigating the weighting of previous averages by β.

Figure 3.3.2: Assignment of weight through time

Today A =	αD_{t-0}	$+\beta A$					
Yesterday		αD_{t-1}	$+\beta A$				
Day before			αD_{t-2}	$+\beta A$			
Day t-3				αD_{t-3}	$+\beta A$		
Day t-4					αD_{t-4}	$+\beta A$	
Day t-5						αD_{t-5}	$+\beta A$
weight	α	$\alpha\,\beta$	$\alpha\,\beta\beta$	$\alpha\,\beta\beta\beta$	$\alpha\,\beta\beta\beta\beta$	$\alpha\,\beta\beta\beta\beta\beta$	
	β^0	β^1	β^2	β^3	β^4	β^5	
age	0	1	2	3	4	5	

Today's average for tomorrow is calculated from today's datum (t-0) and yesterday's average. In turn, yesterday's average was calculated from the previous day's datum (t-1) and average from the day before (shown in red). So yesterday's datum is weighted twice: first by β in multiplying βA today, then by α in computing yesterday's average, αD_{t-1}. As one moves back through time the forecast has been based on the previous day's datum and average, while the weight assigned to past data diminishes, β being a fraction less than 1.

As data become older, so their inclusion in the forecast for today is discounted in proportion to their age. The weight given to past data diminishes *exponentially* (β^{age}), age appearing as an exponent. The new formula, therefore, assigns weight to past data according to the number of times they have been used in previous averages.

3.4 From moving to smoothing

Brown calls this new formula *smoothing* where "...the new smoothed value is equal to the previous smoothed value plus a fraction α of the difference between the new observation and the previous smoothed value..." (1963: 101). The weight assigned to each datum is, therefore, a function of how old the data are, and can be represented by expressing the weight given to past data as:

$$Weight(Age) = \alpha \times \beta^{age}, \text{ where today's age is zero.}$$

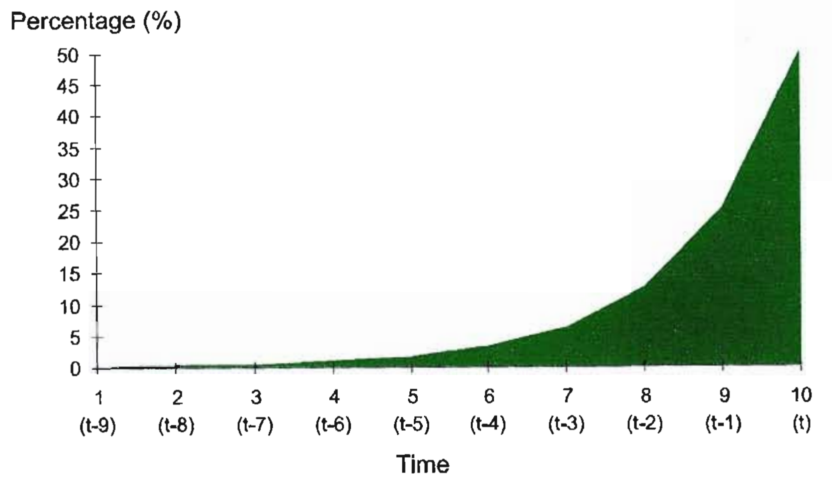
Table 3.4.1: The assignment of weight with α=0.2, 0.5 and 0.8 (β=1-α)

Age	α=0.2 $1/5 \times (4/5)^{age}$	α=0.5 $1/2 \times (1/2)^{age}$	α=0.8 $4/5 \times (1/5)^{age}$
0	0.2	0.5	0.8
1	0.16	0.25	0.16
2	0.128	0.125	0.032
3	0.1024	0.0625	0.0064
4	0.08192	0.03125	0.00128
5	0.065536	0.015625	0.000256
6	0.0524288	0.0078125	0.0000512
7	0.0419430	0.0039062	0.0000102
8	0.0335544	0.0019531	0.0000020
9	0.0268435	0.0009765	0.0000004
Sum	0.8926257	0.9990233	0.9999998

In Table 3.4.1, the weight given to each day's datum is shown, today's age being equal to zero. Where more weight is given to recent data, for example, alpha (α) equal to 0.8, less weight can be given to past data, although all data do, in principle, receive some weight.

The sum of all weights, shown in the last row, must eventually equal 1 or 100 percent, which is simply $\alpha + \beta$, or all the weight that can be assigned to data. The more weight given to recent data, the more responsive a forecast is to recent changes in the data. This is obvious with α=*0.8* where the weight assigned to past data quickly diminishes as time passes, and nearly 100 percent of the weight is assigned in just 9 data. The distribution of weight assignment for α=*0.5* can be seen in Figure 3.4.1.

Figure 3.4.1: Weight assignment with α=0.5 (β=0.5)



(Source: Brown, 1983: 102)

Figure 3.4.1 shows how "smoothly" smoothing applies weight to past data, where recent data receive more weight than distant data. Unlike moving, recent data are considered more important and included to a greater degree in the forecast for tomorrow, an idea truly consistent with the purpose of weighting data by their age.

3.5 Smoothing *versus* moving

"...Smoothing produces an average in which past observations are geometrically discounted according to their age. A moving average weights the M most recent observations each $1/M$, and all earlier observations have weight zero..." (Brown, 1963: 106). This proportional assignment of weight by smoothing *seems* intrinsically more meaningful than with moving. Determining whether this is a fact, requires a comparison of the two methods.

Before this can be carried out, a way must be found for comparing the *magnitude* of α and M, irrespective of their ways of assigning weight (equally or progressively). This can be done by having the "...same average age of the data..." (Brown, 1963: 108). The *average age* for data in a moving period of 5, for example, can be computed by adding each age $0 + 1 + 2 + 3 + 4 = 10$, and dividing by the number of data in the moving period (Brown, 1963: 107), $10 \div 5 = 2$.

As for computing the average age, one may note that adding the ages of the data in a moving period involves the sum of an arithmetic series which starts at age zero, today's datum. The general formula for the sum of an arithmetic series is $(n+1) \times n/2$, where n is the number of data in the series. It is one less than the moving period because the series starts at zero.

$$\text{Sum of ages} = (n+1) \times n/2 = (4+1) \times 4/2 = 10$$

The average age is then computed by the sum divided by the moving period:

$$\text{Average age} = \text{Sum of ages} \div M$$

In order to derive a formula for computing the average age, $M-1$ can be substituted for n in the sum of a series formula: $[(M-1)+1] \times M/2$. Therefore:

$$\text{Sum of time} = \frac{M\,(M-1)}{2} \qquad \text{Average} = \frac{M-1}{2}$$

The sum is divided by M to compute the average. The average age is $(M-1)/2 = 4/2 = 2$. On the average, data in a moving period of 5 are 2 days old.

When the data are *equally* weighted, the formula $(M-1)/2$ is simple enough. However, when the data are *unequally* weighted, one must compute the sum of each age, each one multiplied by its weight. Table 3.5.1 has been compiled for this purpose, with reference to Figure 3.4.1 for the weights assigned to each datum according to age.

Table 3.5.1: The average age for smoothing

Age	Weight = $\alpha \times \beta^{age}$	Age × weight
0	$\alpha \times \beta^0$	0α
1	$\alpha \times \beta^1$	$1\alpha\beta$
2	$\alpha \times \beta^2$	$2\alpha\beta^2$
3	$\alpha \times \beta^3$	$3\alpha\beta^3$
4	$\alpha \times \beta^4$	$4\alpha\beta^4$
		$\alpha \sum k\beta^k$ (k=0 to ∞)

The term for the sum of all ages multiplied by their weight can be seen as $\alpha\sum k\beta^k$, where k is a series from zero to infinity. This sum approaches "β/α" (Brown, 1963: 107). For smoothed data, where $\beta=(1-\alpha)$, the formula for the average age then simplifies to $(1-\alpha)/\alpha$. By equating the formulae for the average ages of moving and smoothing, and solving, one can determine α with a response comparable to M.

$$\frac{M-1}{2} = \frac{1-\alpha}{\alpha}$$

$$\alpha = 2/(M+1)$$

Therefore, a moving period of 5 will produce results comparable with those produced with a smoothing constant of $\alpha = 2/(5+1) = 2/6 = 33\%$ (1/3). Table 3.5.2 lists a number of moving periods with their equivalent smoothing constants.

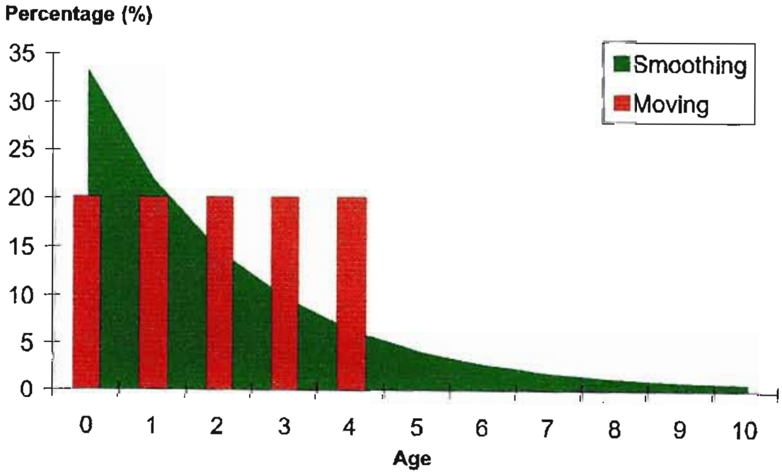
Table 3.5.2: Comparable moving periods and smoothing weights
for the one most recent datum

M	1/M	α
1	1	1
2	0.5	0.67
3	0.33	0.5
4	0.25	0.4
5	0.2	0.33

(Source: Brown, 1963: 108)

The comparison of how differently moving and smoothing assign weight to past data, is most obvious when seen graphically, shown in Figure 3.5.1. A moving period of 5 is used (α=0.33), over a period of ten days (age=0 to 9).

Figure 3.5.1: Weights assigned by moving and smoothing



(Source: Brown, 1963: 102)

The random data used in section 3.2, 3 9 5 7 9 2 7 8 8 2, will be used in Table 3.5.3 to compare moving to smoothing, with moving periods of 1 to 5, and their equivalent smoothing constants, expressed as a percentage.

Table 3.5.3: Comparing moving to smoothing

Moving		Smoothing	
M	Sigma	Sigma	α (%)
1	4.36	4.36	100
2	3.48	3.44	66.67
3	3.24	3.40	50
4	3.27	3.49	40
5	3.18	3.39	33.33

Sigma for both moving and smoothing are of similar magnitudes, which shows that the average age formulae were derived correctly. Where M=1 and α=100%, sigma is identical for both moving and smoothing, because the forecast for tomorrow is the datum from today in both cases. The shaded areas in Table 3.5.3 highlight where moving was better than smoothing, as indicated by lower values of sigma.

Moving was slightly better on three occasions, specifically when M was large, that is, less responsive. However, the data used in Table 3.5.3 are random (taken from Trueman, 1981: Appendix E), and moving was expected to be relatively less responsive to new data than smoothing (Figure 3.5.1). The fact that smoothing is *worse* than moving with random data, paradoxically proves the point, namely that smoothing is *more* responsive than moving. Conversely, it can be concluded that either smoothing is worse than moving, or else the best response is a *slow* response.

3.6 Summary

The least squares methodology has been extended so that the best forecast can be judged by the criterion of minimising the standard deviation or sigma. While sigma is of the same magnitude as the mean absolute deviation, it tends to be somewhat larger, because large errors have been weighted more heavily than small ones.

Furthermore, smoothing is better than moving for three reasons:

Firstly, smoothing is inherently consistent, assigning weights proportionally to a datum's age, whereas moving assigns weight abruptly;

Secondly, it is more responsive to recent data, less responsive to distant data, but in principle gives weight to all data;

Thirdly, it is faster to compute, requiring the storage of only the average and one datum.

CHAPTER 4

Computational aspects of fitting a given model

4.1 Introduction

The application of time series analysis to manufacturing will require the timeous solution of many complicated calculations, such as those explained in previous chapters. Fitting a model with the established equations and formulae can hardly be performed by hand, because of the large number and magnitude of real data. Unfortunately, given the present formulae, these calculations cannot be performed by a computer either.

For example, although two simple formulae were derived to compute the origin and slope in one step, simultaneous solution of equations cannot be performed by a computer, if it were to follow the intuitive so-called "substitution method". Rather, a fast and reliable mechanical procedure is desired that does not require any intuition. *Matrix multiplication* provides one method for matching the mechanical operations a computer can do, with intuitive mathematics.

4.2 Matrix multiplication in general

As for speed and simplicity of computation, one can take advantage of matrix multiplication to obtain the sums required by the least squares method. This is a standard mathematical method requiring minimal human intervention, which makes it suitable for the computer. It involves routinely multiplying the cells in the *columns* of one matrix by the corresponding cells in the *rows* of the other.

For example, given the row and column vectors from two matrices X and Y, the procedure for multiplication is as follows:

$$X = 4 \quad 6 \quad 9 \quad 5 \quad 7 \qquad\qquad Y = \begin{matrix} 6 \\ 12 \\ 3 \\ 12 \\ 6 \end{matrix}$$

$$\begin{aligned} X \times Y &= (4 \times 6) + (6 \times 12) + (9 \times 3) + (5 \times 12) + (7 \times 6) \\ &= 24 \;+\; 72 \;+\; 27 \;+\; 60 \;+\; 42 \quad = \mathbf{225} \end{aligned}$$

(Source: Kemeny, Schleifer, Snell and Thompson, 1962: 235)

Each row element in the matrix X is multiplied by the corresponding column element in matrix Y. This process is repeated for each row and column in the two matrices. Therefore the number of rows and cells in one matrix must be equal to the number of columns and cells in the other matrix, resulting in a square product matrix.

4.3 Matrix multiplication for the covariance matrix

Table 2.4.7 is repeated in Table 4.3.1, with a view to matrix orientation. The simple sums required were those for the time vector, the data vector, the data vector multiplied by the time vector, and the time vector multiplied by itself. These sums for each row are totalled in the last column.

Table 4.3.1: Data used for simultaneous equation solution

t	1	2	3	4	5	15
D	10	20	30	50	40	150
t × D	10	40	90	200	200	540
t^2	1	4	9	16	25	55

The sums of 540 and 55 in the lower right part of the table, those involving the product of two rows, that is, the sum of the time vector multiplied by the data vector, and the sum of the time vector multiplied by itself, can be obtained by matrix multiplication. However, paradoxically, the simple sums of *time* and *data*, in the upper part of the table, cannot be computed in this manner.

As they stand, these totals require simple addition, not multiplication. One could of course write a separate routine for these at the expense of having two routines, one for the simple sums and the other for those involving the products. However, *all* sums *can* be obtained by one simple loop if one remembers that a number remains unchanged if it is multiplied by 1. In other words, by introducing a "dummy" vector containing only the number 1, the so-called *unity* vector, a standard multiplication and summing routine can be used for all the sums required.

By adding the unity vector to the basic data matrix, the simple sums for time and data are computed by multiplying each vector by the unity vector. Using the simple data 50 40 30 20 10, this is shown in Table 4.3.2. A "clean" series is used, that is, a straight forward trend, so that the multiplications can be simply illustrated because one is using perfect data.

Table 4.3.2: Data matrix

Unity	1	1	1	1	1
Cause (time)	1	2	3	4	5
Effect (data)	50	40	30	20	10

Strictly speaking, multiplying the matrix by itself requires it to be rewritten with each row arranged in a corresponding column, as shown in Table 4.3.3.

Table 4.3.3: Data matrices to be multiplied

Matrix A					
Unity	1	1	1	1	1
Cause	1	2	3	4	5
Effect	50	40	30	20	10

Matrix B		
Unity	Cause	Effect
1	1	50
1	2	40
1	3	30
1	4	20
1	5	10

Now that the data are arranged in two matrices A and B, the standard routine explained in section 4.2 can be applied to obtain all the sums required.

Specifically, multiplying the unity vectors by each other (shaded), yields the *number* of data, a sum previously obtained through counting on one's fingers. Now it is obtained by summing, $1 \times 1 + 1 \times 1 + 1 \times 1 + \ldots$, N times, which naturally yields N the number of data. This may seem a complicated method for counting one's data, but it has the advantage of permitting use of one standard routine for *all* the sums required.

This applies similarly to the simple sum of the data. On the other hand, multiplying the time vector by itself, then the data vector by the time vector, involves the same *true* multiplications previously done by hand. A complete matrix multiplication involves also multiplying the data column by the data row. This naturally yields $D_1 \times D_1 + D_2 \times D_2 + D_3 \times D_3 + \ldots$, the sum of the squared data.

In short, matrix multiplication, including the unity vector, obtains not only all the sums previously computed by hand, but also the sum of the squared data. This will subsequently be used to compute the residual variance (the sum of the squared errors), *automatically*, while solving the simultaneous equations. This sum was previously computed after the fact, in a separate operation.

The procedure for counting the number of data, the sum of the unity vectors multiplied by each other, is shown in red in Table 4.3.4. The results of the other multiplications and additions appear in the matrix on the far right.

Table 4.3.4: Covariance matrix computation

Unity	1	1	1	1	1
Cause	1	2	3	4	5
Effect	50	40	30	20	10

×

Unity	Cause	Effect
1	1	50
1	2	40
1	3	30
1	4	20
1	5	10

=

	Unity	Cause	Effect
Unity	5	15	150
Cause	15	55	350
Effect	150	350	5500

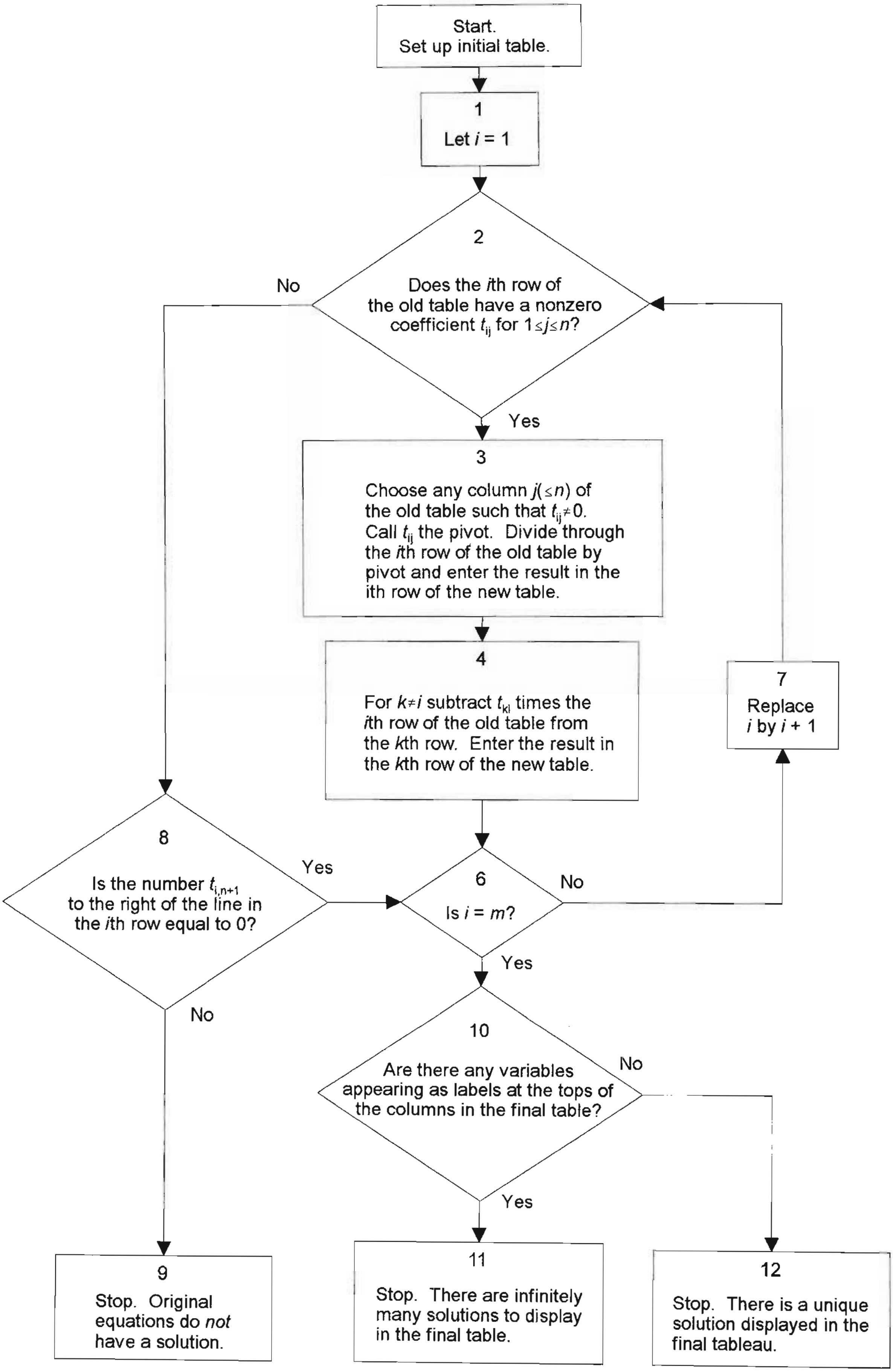
The matrix finally obtained in Table 4.3.4, is commonly known as the *covariance matrix*. Its essential *co*-variances are those between the cause and effect vectors. The so-called co-variances involving the unity vector yields the number of data and their sums, while the "co"-variance for the data themselves yields the variance properly speaking.

4.4 Solution of the covariance matrix

As one used the standard method of matrix multiplication including the unity vector, to obtain the covariance matrix, so a standard method can be used to solve it. It is called matrix inversion.

This method is explained in Kemeny, Schleifer, Snell and Thompson (1972), by means of a flow chart. This flow chart is reproduced in Figure 4.4.1, with certain typographical changes. The flow chart's *t*-variable refers to "tableau" or table, here the covariance matrix of *n* rows and *m* columns. Step five has been omitted because it involves labelling the covariance matrix, known in this case to always be symmetrical, and having the labels *Unity*, *Cause* and *Effect*.

Figure 4.4.1: Flow chart for matrix inversion procedure



(Source Kemeny *et al*, 1972: 193)

In other words, for symmetrical matrix inversion...

1.. *Identify the PIVOT, normally the next item on the diagonal. If done, new table contains identity matrix and solution.*

2.. *Transfer old table row and column headings to new table.*

3.. *NEW PIVOT ROW = OLD PIVOT ROW / PIVOT.*

4.. *In the PIVOT COLUMN of the old table, take the next item other than the PIVOT. Call it CLEAR. If none, go to 1.*

5.. *NEW CLEAR ROW = OLD CLEAR ROW - CLEAR × NEW PIVOT ROW.*

6.. *Go to 4.*

(Source: Adapted from Diegel, 1987: 3)

These steps can be applied to the covariance matrix for the data 50 40 30 20 10 shown in Table 4.3.4, repeated in Table 4.4.1, with the results from the first pivot. The superscript numbers identify which of the above six steps is being performed. Note that the data, obviously, have a negative slope of -10, and that this slope persists perfectly.

Table 4.4.1: Covariance matrix and first pivot

	Unity	Cause	Effect
Unity	5	15	150
Cause	15	55	350
Effect	150	350	5500

	Unity^2	Cause^2	Effect^2
Unity^2	1^3	3^3	30^3
Cause^2	0^5	10^5	-100^5
Effect^2	0^5	-100^5	1000^5

PIVOT = 5^1, CLEAR = 15^4, 150^4

After the second pivot, the solution (inverse) for the covariance matrix can be seen in bold in the right hand table in Table 4.4.2. The sum of the squared errors, zero, appears below the best origin (60) and slope (-10) respectively.

Table 4.4.2: Second pivot and solution

	Unity	Cause	Effect
Unity	1	3	30
Cause	0	10	-100
Effect	0	-100	1000

	$Unity^2$	$Cause^2$	$Effect^2$
$Unity^2$	1	0^5	$\mathbf{60^5}$
$Cause^2$	0	1^3	$\mathbf{-10^3}$
$Effect^2$	0	0^5	$\mathbf{0^5}$

PIVOT = 10^1, CLEAR = -100^4, 3^4

Intuitively, one can see that the computed slope and origin are correct because the data are so simple: one has a constant rate of change, so one obtains a perfect fit. The residual variance or sum of the squared errors is, therefore, zero.

Fitting a trend line to irregular data, for example, 10 20 30 **50 40**, with 40 and 50 reversed, should result in an imperfect fit. The data matrix for this multiplication is shown in Table 4.4.3,.

Table 4.4.3: Data matrix of non-linear data

Unity	1	1	1	1	1
Cause	1	2	3	4	5
Effect	10	20	30	50	40

The multiplication and solution of Table 4.4.3 results in the covariance matrix and its inverse, as shown in Table 4.4.4.

Table 4.4.4: Covariance matrix and its inverse

	Unity	Cause	Effect
Unity	5	15	150
Cause	15	55	540
Effect	150	540	5500

	Unity	Cause	Effect
Unity	1	0	**3**
Cause	0	1	**9**
Effect	0	0	**190**

The origin, slope and residual variance are again shown in bold in Table 4.4.4. The residual variance is 190, highlighted in bold, which is the same as was computed in Table 2.4.5 by hand, using the same data. The origin (3) and slope (9), also in bold, are the same as those computed in section 2.4. Conversely, the residual variance for an imperfect fit is obviously not zero.

4.5 Summary

Matrix multiplication and inversion provides a mechanical and fast method for fitting the best linear algebraic model to *any* data, perfect or otherwise. Matrix multiplication requires no human intervention, only computing power, thus lending itself directly to computer application. It also yields not only the best fitting coefficients for any set of data, but also the residual variance (the sum of the squared errors), automatically.

CHAPTER 5

Special case of fitting time series

5.1 Introduction

Chapter 4 explored the fitting of the simple trend line, by means of mechanical computation. Naturally, the method must be extended to more complex models, including curves and periodic/trigonometric series, and, in principle, to a practically unlimited number of data.

However, before simply extending matrix multiplication to solve higher order models in a routine fashion, one should consider the question of data arrangement and storage. In addition, one must consider certain problems of precision, and obviously, avoid division by zero.

5.2 Data storage and limitations

The standard method of matrix multiplication requires the storage of both matrices, A and B, so that each row can be multiplied by its corresponding column. If these matrices were to include *all* the data currently available, as is presently required, their size would slow down computations. Indeed, some programming languages place quite narrow limits on the largest allowable matrix. Complete storage may be impractical if not impossible, certainly if both A and B matrices are stored explicitly. Therefore, one should like to find a way of overcoming the limitations of programming languages, whilst being able to change the raw data vector without having to reconstitute the data matrix.

One approach results from studying the sequence of operation, first for the matrices in their original order ($A \times B$), then after exchanging their positions ($B \times A$). Table 5.2.1 shows the matrices used in the previous multiplication, where the shaded row in matrix A was multiplied by the shaded column in matrix B.

Table 5.2.1: Traditional order of matrices to be multiplied

Matrix A					
Unity	1	1	1	1	1
Cause	1	2	3	4	5
Effect	50	40	30	20	10

Matrix B		
Unity	Cause	Effect
1	1	50
1	2	40
1	3	30
1	4	20
1	5	10

Since both matrices are numerically identical, one should obtain the same result after reversing the order of multiplication. Table 5.2.2 shows multiplication of matrix *B* (shaded row) by matrix *A* (shaded column).

Table 5.2.2: Multiplication after reversing the order of A and B

Matrix B		
Unity	Cause	Effect
1	1	50
1	2	40
1	3	30
1	4	20
1	5	10

Matrix A					
Unity	1	1	1	1	1
Cause	1	2	3	4	5
Effect	50	40	30	20	10

Obviously, the results of multiplying matrix *B* by *A* (row by column) in Table 5.2.2, are the same as achieved in Table 5.2.1, and previously in Table 4.3.4. All the multiplications are repeated, although in a different order.

Swapping the order of multiplication of the matrices may seem trivial, but it has enormous potential for improving the efficiency of data storage. As each row of matrix *B* is multiplied by the corresponding column in matrix *A*, only the one current datum, one time value and the unity number are involved. Previously, multiplication was conducted on a vector-by-vector approach, which required the full storage of all the vectors. This is impractical with very many data, such as data from a manufacturing firm. All that is required now is the current vector, from both data matrices, which contains the one current datum, current time, and the unity number.

As a second step, one notes that in the special case of time series analysis, the matrices B and A are numerically identical, differing only in their arrangement by column *versus* row. Therefore, the second matrix, or the column by which each row is multiplied, can be eliminated by multiplying the first matrix by itself. In other words, *one really needs to store only one matrix, not two.*

Inspection of the shaded areas in Table 5.2.2 reveals that at any one time, *only one row of the matrix is operative*. That is, one multiplies (1 1 50) by (1 1 50), then (1 2 40) by (1 2 40), each independently of the other. If only one row is operative, then one need *only store this one row*. This means a double advantage:

- unlimited number of data;
- new rows can be added as they become available, without actually storing the previous ones. Only the sums-to-date are stored in the covariance matrix.

No further storage space is required by introducing new data, because, at any point in time, only today's datum, time and unity values are stored. On the other hand, all the totals for the covariance matrix have to be stored, but the covariance matrix is comparatively small: its size depends on the number of *coefficients*, not on the number of *data*.

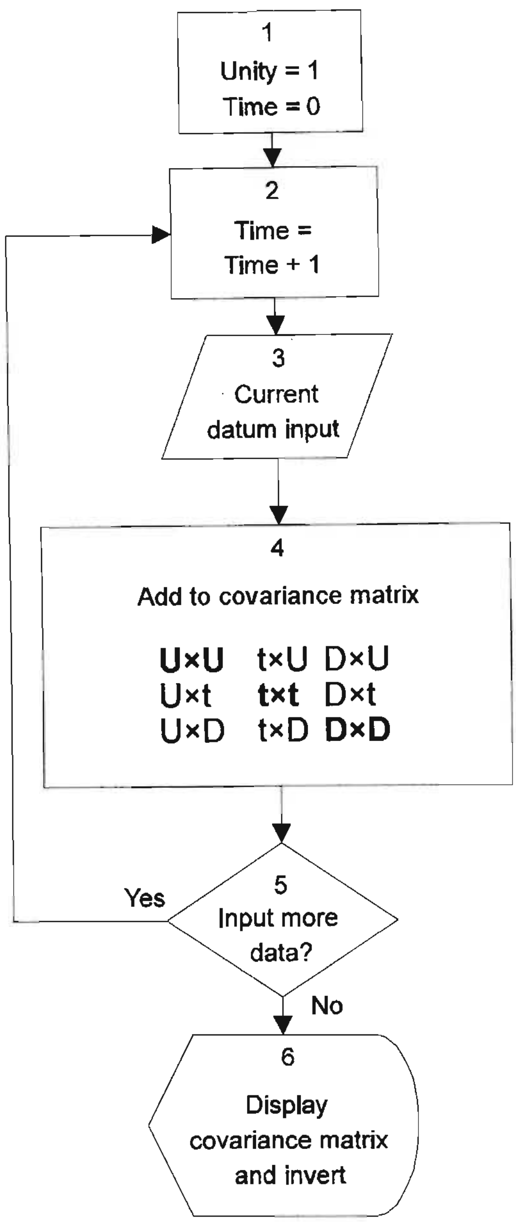
5.3 Vector generation

The unity vector is always 1, so it can be stored as a constant rather than a vector. Time increases by 1 as each new datum is included in the multiplication for the covariance matrix. So it can be *generated* as one goes along, without it having to be stored in a vector at all. In other words, all time-related data can be generated by the multiplication process itself. New data can be included in the computation of the covariance matrix because multiplication progresses with time, as opposed to the vector-by-vector approach in the traditional method.

By simply continuing the multiplication process, increasing time by 1 from day to day, and including the new data, the matrix multiplication can be extended to any number of data, without having to store them at all. One may want to store the data for historical/archival purposes, but they need not be stored for the sake of obtaining the covariance matrix. Only the covariance matrix itself will have to be stored in its entirety, so that it can be solved.

The program SHOWLOOP.PAS is listed in Appendix 2. It is a Pascal program that shows the loop for matrix multiplication with, and without, complete matrix storage. The significant section to compute the covariance table without storing all data, is shown in the form of a flow chart in Figure 5.3.1.

Figure 5.3.1: Flow chart for SHOWLOOP.PAS program



The variable UNITY is set constant to one, and TIME reset to zero.

TIME is increased by 1 each time the loop is repeated. The time vector is thus *generated* by the repetition of the loop.

The current datum can be read by the computer from disk or be entered by the user. In either case, only the *current* datum is involved at any point in time.

The multiplications added to the totals in the covariance matrix, are computed with only the current datum being stored. TIME is generated and UNITY is held constant.

If there are more data to be included in the covariance matrix, the loop is repeated.

If not, the covariance matrix can be displayed. The inverse must then be computed, as was shown in Figure 4.4.1, to solve the matrix.

There are now only three items used in the multiplication and summing for the covariance matrix. Moreover, it is composed of two symmetrical halves, the top-right triangle (shaded) being the same as the lower-left (also shaded). This is shown in Table 5.3.1 with the identical halves shaded.

Table 5.3.1: Covariance matrix

	Unity	Cause	Effect
Unity	5	15	150
Cause	15	55	350
Effect	150	350	5500

Calculations thus actually need to be carried out for only one half, and this is then copied to the other, reducing the number of calculations by half. Looking at the covariance matrix, the unity value in multiplications and sums can be replaced by time with an exponent of zero ($t^0=1$), and cause with time to the power of 1 (t^1). The covariance matrix from Table 5.3.1 is revised in Table 5.3.2.

Table 5.3.2: Revised covariance matrix

	Time^0	Time^1	Data
Time^0	5	15	150
Time^1	15	55	350
Data	150	350	5500

The trend or linear model has two coefficients, that is, the origin and the slope, both of which, along with the residual variance, are shown in the right hand column of the inverted covariance matrix. However, referring to section 2.5, a new column is added for every coefficient to be added to the formula, and a new row for time, with the corresponding power. Thus far, the example has been restricted to fitting a trend to a set of simple data. Computing a solution with a curved model, time squared, for example, will merely require an additional column and row in the covariance matrix.

5.4 Fitting higher order models

The data 6 10 16 24 34, from Table 2.5.1, were used to explain the fitting of a curve/trinomial. Fitting these data with matrix/row multiplication, including the value of time squared in the multiplication with the other operative values, results in the covariance matrix shown in Table 5.4.1. The data used to compute the matrix, not stored but available as time unfolds, are also shown in Table 5.4.1.

Table 5.4.1: Curve covariance matrix

$Time^0$	$Time^1$	$Time^2$	Data
1	1	1	6
1	2	4	10
1	3	9	16
1	4	16	24
1	5	25	34

	$Time^0$	$Time^1$	$Time^2$	Data
$Time^0$	5	15	55	90
$Time^1$		55	225	340
$Time^2$			979	1424
Data				2124

The inverse of this matrix is shown in Table 5.4.2.

Table 5.4.2: Curve solution matrix

	$Time^0$	$Time^1$	$Time^2$	Data
$Time^0$	1	0	0	**4**
$Time^1$	0	1	0	**1**
$Time^2$	0	0	0	**1**
Data	0	0	0	**0**

The solution is $4 + 1 \times t + 1 \times t^2$, with no residual error, shown in bold in Table 5.4.2. Substituting these coefficients back into the fit proves the accuracy of the solution:

$$4 + 1 + 1 = 6, 4 + 2 + 4 = 10, 4 + 3 + 9 = 16, \ldots$$

The matrix multiplication methodology can, therefore, be extended to fit any polynomial model by simply including the extra value for time in the multiplications. In addition to fitting polynomial models, periodic or seasonal data can also be fitted, simply by including the corresponding seasonal terms in the computation.

To illustrate the fitting of seasonal terms, one must first review the components of such a model. A simple *repetitive wave* has at least one polynomial coefficient for the level, and one pair of sine and cosine coefficients to control the wave pattern. Listed below are definitions relevant to the fitting of repetitive data. More precisely, one of these determines the ups and downs of a wave pattern, the other one the location of the peaks:

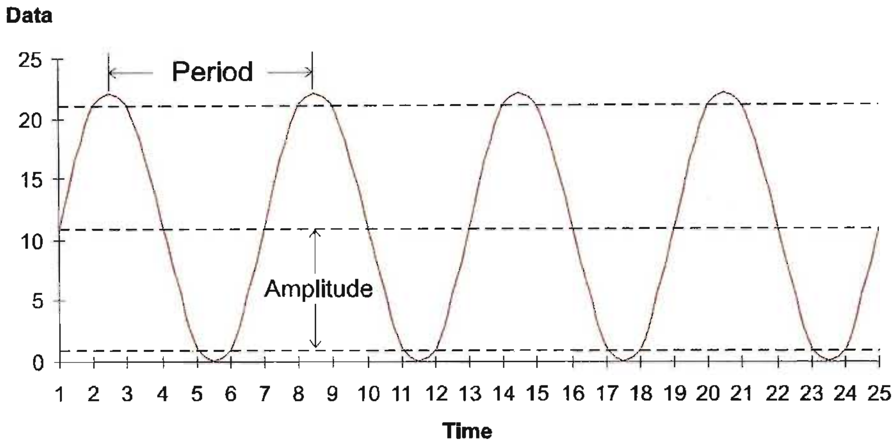
Periodicity: *The distance between two corresponding points on a graph, for example, the two highest points.*

Amplitude: *The distance from the highest to the lowest point on the curve, divided by 2, that is, the curve's maximum variation from the mean.*

Phase: *Location of the peaks (or valleys), that is, the point at which the data change from increasing to decreasing, or vice-versa.*

The sine term determines the amplitude of the peaks, while the cosine term controls the location of the peaks and troughs, shifting the phasing to the left or to the right. The periodicity of the data is used in computing both coefficients. A set of 25 data with a simple wave are 11, 21, 21, 11, 1, 1, 11, 21, 21, 11, 1, 1,.... They can be seen in Figure 5.4.1.

Figure 5.4.1: Graph of a simple wave (6-point)



The first peak is located at time 3, in Figure 5.4.1. The periodicity for the data is visually 6, that is, *8.5 - 2.5 = 6*, the distance between the location of any two peaks. This means that each wave cycle spans a time period of 6. The phase angle is calculated from 360 degrees divided by the visually identified 6-point period. The sine and cosine values are computed from this angle, for example:

$$\sin(360/6 \times 4) = -0.866 \text{ and } \cos(360/6 \times 4) = -0.5 \text{ for } t=4.$$

As the value of time changes, so do the values for sine and cosine depending on whether the wave is going up or down.

The coefficients of the sine and cosine terms in the model determine by how much the wave goes up and down, the amplitude, and these must be solved to fit the periodic data. This can be done by computing a covariance matrix with columns and row for Time^0, Sin and Cosine values. The values required for these multiplications are Time^0 (1), sine at time t (sin (360/period × t)), cosine at time t (cos (360 / period × time)), and the current datum. This is shown for the first 4 of 25 data in Table 5.4.3, along with the covariance matrix.

Table 5.4.3: Wave covariance matrix

$Time^0$	Sin	Cosin	Data
1	0.866	0.5	11
1	0.866	-0.5	21
1	0	-1	21
1	-0.866	-0.5	11

	$Time^0$	Sin	Cosin	Data
$Time^0$	25	0.87	0.5	275
Sin		12.75	0.43	78.81
Cosin			12.25	-114.5
Data				4625

The inverse of the covariance matrix for these data computes the coefficients to be 11, 5.77 and -10, with a residual error of zero. This means that the data are perfectly fitted by the formula:

$$11 + 5.77 \sin (60 \times t) - 10 \cos in (60 \times t)$$

So, in principle, the same matrix/row multiplications can be used to fit the data, using the *least squares methodology*, for periodic models as well as polynomial ones.

In Figure 5.4.1, the amplitude of the wave appeared to be 10, the highest value being 21, less the mean of 11. However, the curved nature of the graph means that the fitted peak is not the highest observed value, but slightly higher, the true amplitude being computed by the sine and cosine coefficients (R and S) in the formula: Amplitude = $\sqrt{}R^2 + S^2$. Therefore, $\sqrt{[(5.77)^2 + (-10)^2]}$, equals 11.55 and not 10.

5.5 Determining the optimal periodicity

Periodicity can often be visually determined, as in the case of the simple data in Figure 5.4.1. For genuine data, periods may not be obvious when "looking" at the data. Furthermore, automation and application to the computer requires a computational routine that is both accurate and objective for any data and for several periods. Indeed, it must also determine whether a given periodicity is strong enough to warrant fitting a wave.

Autocorrelation is a useful tool in determining such a periodicity. It is the correlation of a data series to itself, "auto" Greek for "self", rather than to some other series or cause. In turn, one determines the periodicity of the data series by *lagging* them: this means that in a given set of data, any given day is compared to the previous day throughout the series, then to one two days previously, and so on. The data series without any lag, the raw data as they stand, is called the *base* series, to which other series are compared. Indeed, a lag of zero relates the data to themselves as they stand, obviously resulting in an autocorrelation of 100 percent.

By increasing the lag of a series, the phasing is being shifted to the right. This changes the *coefficient* of autocorrelation, unless the data are all equal. This particular correlation is computed as the covariance of the base and lagged series, divided by the variance. The best periodicity, if any, is determined by the lag with the greatest correlation. The resulting vector of correlations, one for each lag, is known as the *autocorrelation function*. The number of lags it takes for the data to fall into phase, defines the maximum periodicity. Table 5.5.1 shows the data from Figure 5.4.1 lagged by 1 to 6 steps.

Table 5.5.1: Lagging a data series (25 data)

time:	1	2	3	4	5	6	7	8	9	10	11	12
0-lag:	11	21	21	11	1	1	11	21	21	11	1	1
1-lag:	21	21	11	1	1	11	21	21	11	1	1	11
2-lag:	21	11	1	1	11	21	21	11	1	1	11	21
3-lag:	11	1	1	11	21	21	11	1	1	11	21	21
4-lag:	1	1	11	21	21	11	1	1	11	21	21	11
5-lag:	1	11	21	21	11	1	1	11	21	21	11	1
6-lag:	11	21	21	11	1	1	11	21	21	11	1	1

The correlation coefficient for each lag can be computed by the following formula:

$$\text{Correlation} = \frac{\text{Covariance}}{\text{Variance}} = \frac{\sum[(\text{Base datum - Base average}) \times (\text{Lag datum - Lag average})]}{\sqrt{[\sum(\text{Base datum - Base average})^2 \times \sum(\text{Lag datum - Lag average})^2]}}$$

Table 5.5.2 shows the autocorrelation function for lags 1 to 8.

Table 5.5.2: Autocorrelation function (%)

Lag	0	1	2	3	4	5	6	7	8
Correlation	100	50	-46.51	-100	-49.31	46.72	100	50	-45.16

There is obviously a 100 percent correlation with a lag of 0, because the data are being compared to themselves. When the data are lagged by 6 steps, there is a perfect (100%) correlation. This means that the periodicity, within the horizon of 8 days, is 6, and that the data may be fitted by a 6-point wave. Similarly, the -100 percent correlation shows a valley at point 3, half the periodicity.

The number of lagged data in autocorrelation is important for discovering the correct periodicity. The end of the first cycle identified by 100 percent correlation (6 steps hence), need not be a *repetitive* wave. It may be a wavelike variation in a series. For an underlying wave process to be correctly identified, the repetitive nature of this first cycle must be established. If a good correlation for a particular periodicity persists more than twice in the given data, it can be considered repetitive. Conversely, to assure that the *first* peak is a true peak and not a chance occurrence, one needs at least one repetition, or, to be on the safe side, at least two. Therefore, for a periodicity of *P*, to be able to compute the location of three peaks *P* steps apart, one needs a minimum of three times as many data as the periodicity *P*. It stands to reason then, that the longest period that can be established, is one third the number of data. Insufficient data (less than three times the period), can mean that the *true* underlying process may not be discovered.

Autocorrelation provides a computationally fast and objective method for determining the optimum period. In turn, secondary periods can be identified once the primary cycle has be removed from the data.

5.6 Summary

Reversing the multiplication order of the matrices from $A \times B$ to $B \times A$ has shown that only one matrix row is actually being multiplied at any one time. This means that the entire data matrix need not be stored to compute the covariance table. Only the current datum must be available in the row to be squared and summed, while time can be generated according to the power of the model to be fitted, and according to its periodicity.

This means that a practically unlimited number of polynomials and waves can be considered in fitting a practically unlimited number of data. In turn, matrix/row multiplication permits efficient utilisation of the computer, always providing the best fit and overcoming the limitations previously preventing large scale application.

Autocorrelation allows the computer to "see" computationally the peaks and valleys of a data series, and by changing its phasing, to determine the optimal periodicity. Again by objective computations, the sine and cosine coefficients can be calculated without human intervention.

CHAPTER 6

Revision of higher order models

6.1 Introduction

Section 3.5 proved analytically and with real data, that smoothing is inherently better than moving. The example in this proof was limited to the revision of the simple average in the light of new data. However, in principle, all models may need some revision to keep them up to date. On the other hand, if the best policy seems to be rapid revision, one may rather want to turn to a higher order model. Brown suggests that "...if you carry out a sequence of trials on some set of actual data and find that you want to use a smoothing constant that is higher than 0.3, check the validity of using the constant model..." (1963: 107).

Revision refers to the coefficients of a model being changed to keep up with recent events in the market. With each new error included in smoothing the average, the forecast level is revised and changed. In a model with many coefficients, one cannot revise one without considering the effect on the others. In addition, with higher order models, one has to consider *updating* as well as revision.

6.2 Updating the trend and higher order models

To begin with, a simple example: if the origin was zero and the coefficient for the slope has been 10 for 5 days, then the forecast for day 6 is 60. If at this point, the slope is to be revised to 15, the forecast for day 6 would be the value at day 5 (50), plus the new slope (15), which is 65. However, if the forecast with the revised slope were made as of day 1, the result would mistakenly be:

Fit (t=6) = A + B × t or *0 + 6 × (15),* which equals 90. To correct this, the old origin of zero must be updated to 50 and time reset to zero, that is: *50 + 1 × (15) = 65*. The revision of the slope of a trend (linear model) cannot be done correctly without updating of the origin to "now".

To update the formula for the simple trend, $Fit = A + B \times t$, the origin A becomes Fit (the previous forecast). The two processes of computing the Fit and then making $A = Fit$, can be simplified by: $A = A + B \times t$. As the simple trend example demonstrated, the slope B itself is not *updated*. Once the origin has been updated, the slope can be *revised* by yesterday's error and the next forecast computed.

Unfortunately, revising higher order models is not so simple. To illustrate this, consider the next higher model, a curve or trinomial. It has coefficients for the origin, slope and the rate of change of the rate of change, for example: $100 - 20 \times t + 3 \times t^2$. This function generates a series, whose first 10 numbers are shown in Table 6.2.1.[1]

Table 6.2.1: Generated data (A=100, B=-20, C=3)

t	1	2	3	4	5	6	7	8	9	10...
D	83	72	67	68	75	88	107	132	163	200...

With the additional term, there is a rate of change of the rate of change, discussed previously in section 2.5. The negative slope results in the data decreasing for the first 3 days, until the extra term becomes large enough to off-set this. The rates of change for the data in Table 6.2.1 are shown in Table 6.2.2.

Table 6.2.2: The changing rate of change

t	1	2	3	4	5	6	7	8	9	10
D	83	72	67	68	75	88	107	132	163	200
ROC	-	-11	-5	1	7	13	19	25	31	37
ROC	-	-	6	6	6	6	6	6	6	6

1 This table, and the other tables in this section and sections 6.2.1 and 6.2.2, as well as certain quotes, are based on unpublished lecture notes produced by Dr Adolf Diegel (1997), for the purpose of studying routine updating.

Unlike for the simple trend, Table 6.2.2 shows that the rate of change is itself changing because of the extra term in the model. "...As a result, as time advances to "now", updating affects not only the origin A, but also the (linear) slope B...". This can be seen by the change in coefficients when the data are shifted one day to the left, shown in Table 6.2.3.

Table 6.2.3: Generated data seen one day later

t	0	1	2	3	4	5	6	7	8	9	10...
D	83	72	67	68	75	88	107	132	163	200	243...

With the observation for day 1 now 72 instead of 83, a standard fit would produce the coefficients: $A=83, B=-14$ and $C=3$. The origin shifts to the excluded datum, the slope weakens from -20 to -14, but the C-coefficient remains constant. The slope weakens because the initial drop from 83 to 72 (-11) is now only 72 to 67 (-5). Table 6.2.4 shows the change in coefficients as up to 6 data are excluded from the beginning of the series.

Table 6.2.4: 3-term model coefficients, depending on start of series

Start at day	Origin A	Slope B	Time squared C
1	100	-20	3
2	83	-14	3
3	72	-8	3
4	67	-2	3
5	68	4	3
6	75	10	3

The linear slope weakens until day 5, when it becomes positive. Indeed, this is the point where the data stop decreasing. "...As for the specific change in the linear rate of change, it is +6, clearly 2×3, or generally, $2\times C$, twice the rate of change of the rate of change...".

In the light of each new datum, the trinomial $A + B \times t + C \times t^2$ is updated by:

$A = A + B \times t + C \times t^2$, forecast for time t becomes origin for next forecast

$B = B + 2 \times C \times t$, slope is updated by $2 \times C \times t$ (twice C becomes the updating interval)

C remains constant

6.2.1 Updating higher order polynomials

Just as the trend and curve were updated, so the same principle of updating can be applied to a quadratic model. For example, the updating for $F(t) = 200 - 30 \times t - 5 \times t^2 + 1 \times t^3$ turns out to be as follows:

$A = A + 1 \times B \times t + C \times t^2 + D \times t^3$, origin updated to forecast for time t.

$B = B + 2 \times C \times t + 3 \times D \times t^2$, slope updated in terms of C and D, and t.

$C = C + 3 \times D \times t$, C updated in terms of D, and t.

D remains constant.

A distinctive pattern seems to be arising for updating polynomial models. The origin is updated by setting it equal to the forecast for today. The terms other than the origin which affect the rate of change, are updated with respect to the rates of change that directly affect them. However, each coefficient except the last, is multiplied by a constant (shown in red) which seems to represent that coefficient's power in the model. That is, A is updated by 1 multiplied by B, then B is updated by 2 multiplied by C and 3 multiplied by D. Similarly, C is updated by 3 multiplied by D. Therefore, C is always multiplied by 2 and D by 3, their powers for time in the model. The last term remains constant.

However, it can be shown for an even higher model, for example,

$F(t) = 200 - 30 \times t - 5 \times t^2 - 2 \times t^3 + 1 \times t^4$, that the updated constants appear as follows:

$$A = A + 1 \times B \times t + C \times t^2 + D \times t^3 + E \times t^4$$

$$B = B + 2 \times C \times t + 3 \times D \times t^2 + 4 \times E \times t^3$$

$$C = C + 3 \times D \times t + 6 \times E \times t^2$$

$$D = D + 4 \times E \times t$$

E is constant.

The constants used in updating the C coefficient, 3 and 6, do not simply reflect the position of the coefficients and no *simple* progression is evident. However, instead of arranging the updating constants side by side (1 2 3 4), one can arrange them in a triangular form, as in Figure 6.2.1.1.

Figure 6.2.1.1: Pascal triangle for updating constants

```
                  1
               1     1
            1     2     1
         1     3     3     1
      1     4     6     4     1
   1     5    10    10     5     1
1     6    15    20    15     6     1
1     7    21    35    35    21     7     1
```

Starting in the first row with 1, and the second row with 1 twice, each value in the subsequent rows is the sum of the two numbers above it, indicated by the black lines in Figure 6.2.1.1. For example, 2 in the third row is $1 + 1$, 3 in the fourth row is $1 + 2$, and 4 and 6 in the fifth row is $1 + 3$ and $3 + 3$ respectively, and so the process continues with each new row. This arrangement of numbers is called a *Pascal triangle*, and it yields the correct constants for updating. Looking at the second slant of numbers in red along the side of the triangle (1 2 3 4): these are the constants for updating B. The third slant (1 3 6) are the constants required for C, those that were problematic because they did not follow a linear progression.

Furthermore, the outer sides of the triangle (1 1 1 1 1), are also the constants used for updating A and E. Although these numbers are always 1, it shows a complete application of the triangle. This triangle can be extended through simple summing to compute the updating coefficients for any polynomial model. Thus the Pascal triangle permits a routine method for updating polynomial models of any power.

6.2.2 Trigonometric models

A trigonometric model uses sine and cosine, as well as polynomial coefficients to compute a repetitive wave pattern according to the periodicity detected in the data. At least the average is required to establish the level from which the wave increases and decreases. A trigonometric model can have as many polynomial terms as any other model, but the perfectly repetitive form of a sine wave means that it has no affect on polynomial slopes. The amplitude does increase the origin by the mean of the wave. As the sine controls the wave amplitude, and the cosine the location in the cycle, both coefficients need to be updated. Furthermore, because they are so inextricably linked, both need to be updated, each in terms of the other. This requires the first coefficient to be stored (Save), before it is updated in the first operation, for use in the second updating operation. This can be seen as follows:

Save = Fit (sin), store sine coefficient

Fit (sin) = Fit (sin) × *t (cos) - Fit (cos)* × *t (sin)*, update sine

Fit (cos) = Save × *t (sin) + Fit (cos)* × *t (cos)*, update cosine

Updating has therefore been extended to include any model, including trigonometric models. Once the coefficients have been brought up to date, they can be revised.

6.3 Smoothing the trend

Section 3.3 showed that when the smoothing factor alpha was greater than zero, the current model is revised in view of the most recent observation, rather than relying completely on the previous unchanged average. Where rapid updating is best, a more powerful (higher order) model might be necessary to represent the true process in the data, because forecasting with the current model results in large errors. That higher model, too, may also benefit from some revision, although not as rapid.

Smoothing the average was a computationally simple procedure requiring two basic steps. These are computing, firstly, the error for today's forecast, *Error = Datum (today) - Average*, and secondly, the new average for tomorrow:

Average (tomorrow) = Average (today) + Alpha × *Error* or *A = A + Alpha* × *(D(now) - A)*.

This procedure can be extended to include the revision of the trend forecast for tomorrow:

$Error = D - A - B$

$A = A + B + Error \times (1 - Beta^2)$ and $B = B + Error \times Alpha^2$

(Source: Diegel, 1973: 1)

Two steps are required for the revision of the trend, and they are more complicated than for the simple average. The first step, as with the average, is that the error must be computed, although the forecast now consists of the trend formula, $A + Bt$. Secondly, each of the two coefficients A and B must both be updated and smoothed, where previously there was only one coefficient. The coefficient A is updated by $A + B$, the forecast for yesterday. The slope is not updated because there is no further rate of change to affect it.

The revision of the trend requires the weighting of both A and B, and not by alpha alone as with the constant model, but by $(1 - Beta^2)$ and $Alpha^2$. Brown (1963: 128-140) shows arithmetically how these equations are derived and why the weights are these particular values. *Double exponential* *smoothing* refers to smoothing each of the trend coefficients in such a way that the combined weight for A and B assigned to the most recent error remains a constant, alpha.

Rewriting the formulae for the trend in vector notation provides a method for calculating the "...weighting operator appropriate for any model..." (Diegel, 1973: 1). In its present form, the trend model or fit includes two coefficients and their corresponding values for time:

$$F(t) = A \times t^0 + B \times t^1$$

This can be separated into two vectors, one for the coefficients and the other for the values of time: $Fit = (A\ B)$ and $Time = (t^0\ t^1)$. Each cell in the fit vector is multiplied by the corresponding cell in the time vector to project the forecast for time t.

The equations previously derived by Brown manually,

$$A = A + B + Error \times W \text{ and } B = B + Error \times X,$$

contain two parts: updating shown in red, and revision shown in blue. A and B are elements of the fit vector. Therefore, the updated and revised fit can be written as:

$$Fit = update \times fit + error \times smooth.$$

Each coefficient in the fit vector is updated by its corresponding cell in the updating vector, containing the Pascal triangle constants. Similarly, it can be seen that the forecast error is multiplied by a smoothing vector, as the error was multiplied by W and then by X. In the case of the trend, the smoothing vector contains the weights $W = (1 - Beta^2)$ and $X = Alpha^2$. What remains in order to expand smoothing to higher order models, is the computation of the correct smoothing vector.

6.4 Smoothing higher order models

Triple exponential smoothing is the revision of the curve, also discussed in Brown (1963: 140-144), but vector multiplication can be extended to include this model and indeed any model, without laborious hand calculations. The fitting vector can be extended at will:

$$F(t) = A \times t^0 + B \times t^1 + C \times t^2 + \dots R \times sin\ (360/period \times t^1) + \dots$$

Similarly, the time vector and Pascal triangle can be extended to include the new coefficients, but the smoothing vector needs to be computed with the particular weights to revise the fit correctly.

When computing the smoothing weights by hand, Brown uses the formula for the sum of an arithmetic series, where "...the basic operation is to find the sums of the products of successive time vectors..." (Diegel, 1973: 2). This time vector contains the values for time, for example, days, as if the forecast was being projected into the future from time zero. This distinctly different *time[0]* vector, contains the numbers (1, 2, 3, 4, ...) to multiply with the fitting coefficients, to the power of the model (exponents) held in the time vector.

Instead of deriving formulae for computing this sum, as done partially by Brown (1963: 140), it can be done by *brute force* multiplication and summing with many iterations. This procedure may seem computationally slow, depending on the number of iterations. Also, that number being limited, the results are not exactly the same as with an analytic formula. The big advantage, however, is that this procedure can be performed by a computer, without the storage of any formulae or variables, and, modern computers being so fast, one can hardly measure the time of computation. Moreover, the procedure can be extended at will.

The following extract from a Pascal program computes the sum of the product of the two *time[0]* vectors, the result appearing in the matrix *T*, the size of which is controlled by the number of coefficients in the model:

`time[1]:= 1;`	First cell in the *time[0]* vector set to 1
`for Iterations:= 1 to X do`	Repeat multiplication and sum *X* times
`begin`	
`time[2]:= Iterations;`	Set next successive time value
`for Line:= 1 to NumCoeff do`	Multiply all values in the first vector by all
`for Column:= 1 to NumCoeff do`	values in second vector, producing a
`T [Column, Line]:= T [Column, Line]`	symmetrical matrix (NumCoeff × NumCoeff)
`+ time [Column] * time [Line];`	Sum all multiplications
`end;`	

This program computes the *un-weighted* T-Matrix, un-weighted because beta (the weight assigned to all past observations) is excluded from the calculations for the sum of time. The matrix requires inversion for its solution.

However, weight being assigned proportional to age in smoothing, one can use the same procedure to compute the *weighted* T-Matrix, by including beta (weight) as time progresses. With each iteration, so the T-Matrix is weighted according to *Beta*, shown in blue.

```
weight:= 1; {or Beta^0}
time[1]:= 1;
for Iterations:= 1 to X do
begin
   time[2]:= Iterations;
   for L:= 1 to NC do
    for C:= 1 to NC do
     T [C,L]:= T [C,L] + time [C] * time [L] * weight;
   weight:= weight * Beta; {or Beta^Iterations}
end;
```

The number of *Iterations*, set by *X* (shown in red), determines how much weight is assigned to past data. The weighted T-Matrix also requires inverting, which reveals the weighting coefficients for the number of coefficients specified.

6.5 Optimal Beta

The number of *Iterations* allowed by *X* determines the extent to which the weight assigned to past data is discounted exponentially with age. With the analytic method, this is assumed to approach infinity. As for the computer, it is not practical to push to infinity, but one can take advantage of the fact that *weights assigned to past data diminish exponentially*, and quickly reach very small numbers. Although it never reaches zero, the numbers are so small as to become insignificant. Deciding on the point where a weight is to be considered insignificant, is a matter of precision and computation time. Nowadays, computation time is relatively fast, but precision is still important.

The higher the alpha, the less weight can be assigned to past forecasts and therefore, the fewer iterations required to reach a very low beta. However, if alpha is small, then considerably more iterations will be required (see Table 3.4.1). If a limit is to be set, it needs to be the minimum weight, the point where it becomes negligible, and not a maximum number for iterations *X*.

As the program listed in section 6.4 reduces weight with each iteration, the loop needs some revision so that the *for Iterations:= 1 to X do begin...end* statement becomes a

repeat...until weight $\leq$ *epsilon,*

epsilon representing the lowest weight to be attained, that is, a weight so small that it is as if the error was discounted to infinity. Diegel (1973: 3) suggests a limit of *1E-8* (0.00000001), based on a study of the accuracy and computational speed of the CDC 6400 computer. This is shown in Table 6.5.1.

Table 6.5.1: Computational speed and precision trade-off

Cut-off point	Precise digits in final result	Relative execution time
1E - 4	2	50%
1E - 6	3	75%
1E - 8	**6**	**100%**
1E -10	7	125%

(Source: Diegel, 1973: 3)

Computers today permit faster mathematical computation, although the execution time still remains in the same proportion to accuracy, as shown in this study. Thus, the power of today's computers permit computations with a smaller *epsilon* in the same time as before. However, this improved accuracy hardly improves the forecast.

6.6 Summary

The calculation of the smoothing vector provides the key for revising a model of any power in the light of new data. Computationally, the revision routine is simple and can easily be adapted to the computer. Updating, however, requires the storage of the Pascal triangle. This need not be a major concern, for a polynomial model with more than six terms is unlikely in a real life situation. Therefore, only the first six lines of the Pascal triangle, shown in Figure 6.2.1.1, need be stored.

CHAPTER 7

Inconsistent inconsistencies of data storage

7.1 Introduction

Most companies keep some kind of record of their sales for accounting and taxation purposes. However, the extent to which these data are conducive to accurate forecasting, depends not only on the data themselves, but the time interval used to record them. There are no regulations to control this, apart from the start of the financial year, and so record keeping varies from firm to firm.

First impressions deemed *Lever Ponds (Pty) Ltd* (Durban) a suitable firm to test the success of forecasting with time series analysis, because of the computerised nature of their operation, and the large number and type of products they manufacture. As a matter of convenience, from this point on, the company Lever Ponds (Pty) Ltd will be referred to as the *client*. The client's manufacturing is divided into two sections, namely *Personal Products* and *Detergents*. Personal Products are items such as shampoos, toothpastes and deodorants, whereas Detergents are washing powders, dish-washing liquids and fabric softeners. In all, their products number close to three hundred. Their various brands are sold nationally in many stores and supermarket chains, and are bought by a great number of South Africans. Indeed, products are so many, so varied and changing so fast, that some computer-based record keeping and forecasting seems inevitable.

7.2 The twelve month year

The general practice of the client, and of many other firms, is to record sales data by calendar months. However, little consideration is given to months being of unequal length, and the effect that this has. In fact, every subsequent month, except for July (31) / August (31) and December (31) / January

(31), has a different length from the previous one. Monthly accumulated sales are, therefore, usually totals for periods of different lengths. This is most obvious with 28 days in February, and 31 in March. The significance is that the sales for February could seem to be 11 percent less than those for March, when in *reality* the average sales per day could well be constant.

Having recorded a sales datum at the end of the month, that is, on day 31 or 30, means that mathematically, the datum is the result of 31 or 30 days of trading. This is not in fact true because, firstly, public holidays, weekends or other reasons for days without trade, means that there were fewer days of sales. Secondly, each day in a given month may have a different volume of sales, yet no consideration is given to sales being higher on certain days than on others. Any underlying pattern occurring, for example, on a weekly basis, is obscured by the large and sometimes unequal time interval.

In addition to the problems of monthly data, is also the norm that "...the accounting months are set up to give thirteen weeks in each quarter, with two months of four weeks each, followed by one with five weeks..." (Brown, 1963: 73). This means that the values for every third month are the accumulation of sales for a full week more than the other months, and should be 25 percent larger, all other things being equal. The interpretation of a 13 week quarter, as distorting as it is, is not consistent throughout industry. Brown refers to this practice as a *4:4:5* week ratio for each quarter.

The client, although also implementing a similar system, have done so in a *5:4:4* week ratio since 1994. Prior to this, there were eight six-week periods and one four-week period per year. The time interval is not only irregular, but changes significantly over the period of study from 1992 until 1996.

As a further note, it is also commonly accepted that there are 52 weeks in a year, but 52 multiplied by 7 is 364 and not 365, as is the case for a non-leap year. Sales for this missing day would be included in the following month's sales, adding one or two extra days to January, depending on whether the previous year was a leap year or not.

7.3 Demand *versus* sales

The data themselves are also problematic because *sales* data, especially when studying manufacturers, are not the same as actual *demand* in the market. The figures represent sales to other suppliers and distributors and not to the actual customers which is the *market* that one is trying to forecast. A system that feeds records of *point of sale* demand should be established to improve the data. It may be logistically difficult for the client to track every sale in every shop, although a substantial amount of data should be available from the larger supermarkets.

Unusual sales activity at the distribution and warehouse points of study can result from a number of events. The announcement of price changes can result in bulk buying by large retailers now, which would not be typical of the market under constant price conditions. Strikes and mechanical upgrading can mean periods of no production, or controlled use of stocks from the warehouse to cover this period of no production. These events can cause random fluctuations in the *routine* activity of a particular product. *Arbitrary adjustments* to correct this are hardly possible because "...it will not be practical to make such detailed investigation of the plausibility of demand records for every item in the inventory..." (Brown, 1967: 115).

The client records, as weekly totals, the sales for each *stock keeping unit* from their four national distribution warehouses. A stock keeping unit is a pack or item as it is sold to retailers, for example, sales for a six pack of product X is recorded separately from the same product in a pack of three. It is necessary to use stock keeping units for forecasting because it prevents one variation of a product interfering with the demand patterns of another. These weekly data are then accumulated into monthly sales. Using a twelve-month/four-quarter year is not conducive to accurate forecasting, and weighting some months more than others is not a true solution to the problem.

Another important factor regarding data storage is how many observations are stored, and available for forecasting. Old data are seldom destroyed or lost, but are archived and placed in storage and not looked at again, because of the amount of disk space and computing power required to keep these outdated records alive. It can then prove difficult and time consuming to extract these data. The important point is whether there is enough data available to allow the accurate discovery of the true underlying process.

7.4 Manipulating the time unit

By definition, a *discrete time series* has a measurable duration between each observation. In order to attempt to forecast with such a series, a set of guidelines were submitted to the client on how their data should best be arranged, which is shown in Appendix 3. This stresses the point that recording the actual date for daily data means than an accurate time vector can be established.

Table 7.4.1 is a practical example of how ignoring missing data affects their fitting. It can be seen that the datum for day 4 is missing, as is noted in the *True time* column, because time jumps from 3 to 5.

Table 7.4.1: Picking up the trend

True time	Sequential number	Data	True fit	$Error^2$	Erroneous fit	$Error^2$
1	1	10	10	0	8	4
2	2	20	20	0	21	1
3	3	30	30	0	34	16
5	4	50	50	0	47	9
6	5	60	60	0	60	0
				0		30

It can be seen visually that the data follow a trend with an origin of zero and slope of 10, even though the datum 40, for day 4, is missing from the *Data* column. Logically, there should be a perfect fit, and it would be obtained if data 50 and 60 were fitted to time 5 and 6, respectively. However, if they were fitted to the *number* of the day on which they were recorded, 4 and 5, the origin is -5 and the slope 13. This results in the sum of the squared errors totalling 30, instead of zero, as is the case when fitting the data with the true time vector. It is obvious how serious this mistake can be, and how easily it can be avoided, if only one is aware of the need to fit data to their *time* and not their *number*.

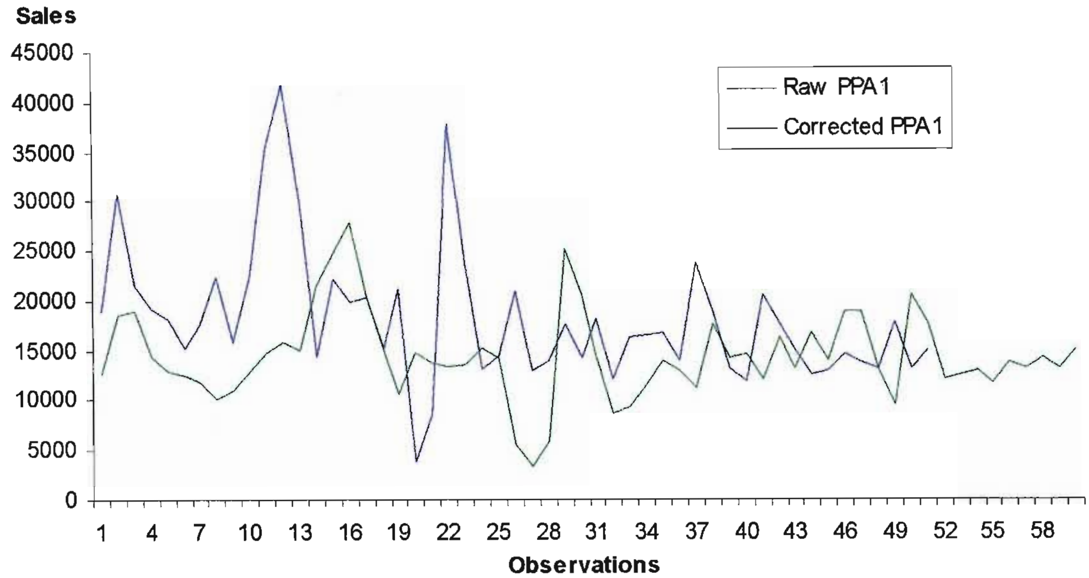
The data from the client were received in hard-copy form, and were totals for the 5:4:4 week months discussed in section 7.2. In order to establish a time unit for the data, each datum was given a time value for the number of days (seven) taken to accumulate it. An example of this is shown in Appendix 4 which shows the actual date for the end of each sales period. Note the *Gap* column, where the first period in each year is 1 day longer than 5 or 6 weeks, because there are not exactly 52 weeks in a year. If the previous year was a leap, this gap is 2 days.

From the values in the *Gap* column, the value of the *time unit* can be established. The time unit is computed from the total time which the observations span, that is, the last date less the first, divided by the number of periods taken to accumulate this time, which is one less than the number of data. This is the average gap between each observation. With sequential time, this value is 1; however, with the time interval for Detergents this value is 30.1739, and with Personal Products, 24.4247. This time unit is then multiplied by the fit to determine the success of the forecast.

The time unit used is not *really* the true unit because it has been generated by manipulating the interval information supplied by the client. Unfortunately, it is the best that can be done with the given information, and improving on this will require considerable commitment and effort from the company to change their current practices.

Figure 7.4.1 graphs the affect of both the 5:4:4 quarter year and the change in the time interval, as compared to a corrected (regular) time unit. In order to visually compare the data for product PPA1 (data listed in Appendix 9), before and after the corrections, it is necessary to plot them on the same graph. Therefore, the x-axis does not represent time but the observation number. The sales values themselves have to be changed in accordance with the time taken to accumulate them, because the x-axis is held constant. As the time unit is not only unequal but uneven, the corrected data were computed from a weekly average, multiplied by a constant 4-week period. This brings the data back to the same magnitude of the given observations, but with equal intervals between them.

Figure 7.4.1: Raw and corrected data for product PPA1



There are now more corrected values, (shown in green) than raw data (blue), but this is similar to the mathematical effect of using the corrected time vector, some observations having taken longer to accumulate than others. Therefore, when using a smaller yet constant interval period, there are more data. It can be seen how exaggerated the initial raw observations are, that is, where the period intervals were 6 weeks long.

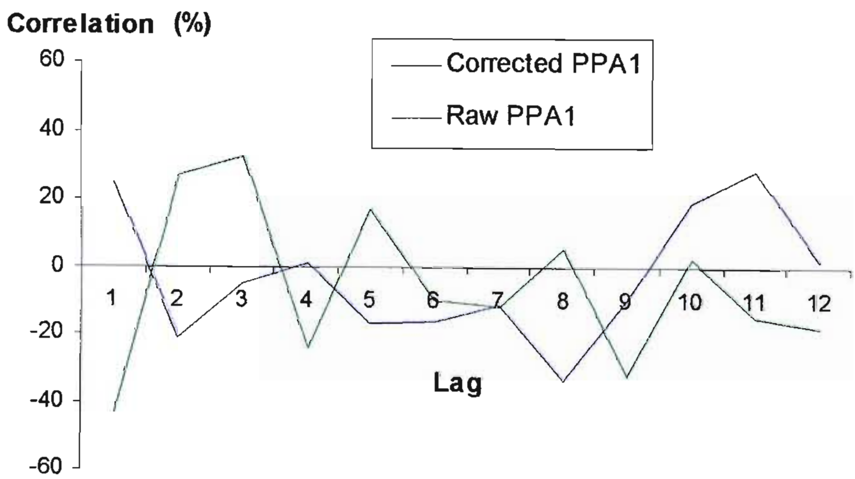
In addition to looking at data levels, it is also important to consider the periodicity of the raw data and the corrected ones. This comparison can be computed using autocorrelation, the results of which are shown in Table 7.4.2. The peaks and troughs identified are shown in colour. A slight periodicity of 3 was detected using a time unit of 24.4247 (corrected values), but there was an even less pronounced periodicity of 4 for a time unit of 1, the data in their original form.

Table 7.4.2: Comparison of autocorrelation functions (Lag=1 to 9)

Lag	1	2	3	4	5	6	7	8	9
1.0000	24.67	-21.46	-5.41	0.86	-16.76	-16.51	-11.10	-33.78	-10.29
24.4247	-43.13	26.74	32.22	-24.05	16.65	-9.89	-12.10	5.18	-32.51

Not only is the periodicity weaker with the uncorrected values, the peak at lag 4 being very small (0.86 ⟨ 1%), but the period length was different, 4 as opposed to 3. This can be seen in Figure 7.4.2, where the autocorrelation function for the corrected values is shown in green, and the uncorrected in blue.

Figure 7.4.2: Autocorrelation functions, raw and corrected



Although there is very little periodicity in the data, it is clear that the autocorrelation functions are quite different as a result of using the restructured time unit. So it is imperative when attempting to fit the client's data, that the correct time vector is used.

7.5 Impact on research methodology

Thus far in this thesis, the *general* problems related to forecasting have been dealt with. As regards the *specific* problems of applying these forecasting techniques to manufacturing, the analysis in this chapter has shown that because of the poor quality of the client's data, they will not permit conclusive research on their own. *The inability of the true data to reveal their underlying process,* is the final *specific* problem identified in section 1.5. To deal with this problem, good quality *generated* data will be used to investigate the recovery of the underlying process. The characteristics or processes in generated data are *known* by definition, and one can objectively establish how accurate fitting was, and how many data were required to obtain this fit. The next task, therefore, is to generate data suitable for experimentation with fitting.

7.6 Summary

"...Garbage in, Garbage out. In other words, an integrated data-processing system is no better than the data it is given to process..." (Brown, 1963: 23). The quality of the data depends not only on how precisely the observations were recorded, but also on how often, and whether the time or date was noted. Observations recorded relative to the lowest time interval, in this case on a daily basis, permit any other transformation into weekly, monthly, or quarterly periods, or whatever other tallying is considered normal for accounting, as well as being suited to forecasting.

The data from the client show that not enough consideration is given to establishing a true time vector. One reason for this is the logistical nightmare of recalibrating existing systems to handle the recording of data on a daily basis. The mathematical advantages are of course greater accuracy of fit, and therefore, forecasting. Any compromise between cost and accuracy, should at least insist on recording the specific number of days taken to accumulate sales.

CHAPTER 8

Generating data with known randomness

8.1 Introduction

It is intuitively obvious that few data are unlikely to show a clear picture of what is happening in the market, as it is only a snap-shot of the grand picture. Responding to misleading suggestions can be worse than not responding at all. For example, data which are random in the long run may have pockets of similar data in the short run. The average will, therefore, be a more successful model in the long run, than following a trend discovered in a few recent data.

The more data available, the more support and justification there could be for a higher order model. The question then arises, how many data are enough? The answer must include drawing the line between over-fitting and discovering the true underlying process.

8.2 Knowing the true process

The algebraic answer to the number of data required to fit a given model, is one datum for the average, two data for the trend, or three for a simple curve. However, this is the case only if the few data analysed do not deviate at all from the true underlying process: they are perfectly representative. *Real* data, not being perfect, mean that a model based on too few observations results in large errors. A strong slope detected may be weakened when considering more data, but otherwise quickly extrapolates into infinity or down to zero. The more polynomial terms in the model with poorly chosen coefficients, the higher the power of time and the greater the potential for large errors because of the higher responsiveness of the model.

With real data, the number of data required for a given model depends on two factors:

- complexity of the model;
- presence of random variation.

Sales and demand data are seldom perfect because so many factors affect their behaviour that they may appear random, if they are not inherently random. In the case of manufacturing, such randomness can be the result of stores offering price cuts on selected items for limited periods, the sales for those items becoming unusually high. Forecasts based on this period of increased sales or demand would clearly be dangerously exaggerated. In-store advertising and promotion can play a similar role in hiding the underlying process one is trying to discover. There are many reasons why people choose one product over another on a particular day. The result is that data generally contain both true trends and cycles, as well as an unpredictable element.

The degree to which this element is present depends largely on the type of product. Fashionable products, such as those manufactured by the client, with many similar competitive products, are likely to have a large number of buyers with very little product loyalty. People's attitudes often appear fickle under such circumstances, and sales are driven largely by advertising campaigns centring on image enhancement, price reduction, or quality. The *true* demand, independent of these promotional peaks, is difficult to establish, certainly if insufficient data are at hand.

Real data being what they are, it is difficult ever to be certain about their true process, and about which part is random. This is not so with generated data because the generation formula is known, and a specific random element can be added. It is then a simple process of experimentation to discover how many data are needed to accurately recover those known coefficients. The findings will be applicable to real data, relative to that random/true process ratio, which can itself also be controlled.

8.3 Generating data with a random element

Any spreadsheet application can generate a set of data from a given formula, with any degree of complexity. For example, below are listed four types of models which are likely to be relevant to business processes. These formulae were used to create sets of 200 data:

Average:	50.5 (data range from 1 to 100)
Trend:	$450 + 10 \times t$ (origin 450 and slope 10)
Curve:	$450 - 10 \times t + 0.1 \times t^2$
Season:	$450 + 100 \sin 72° \times t - 100 \cos 72° \times t$ (5-point wave)

These coefficients were selected so that the data would be of comparable magnitude and manageable for study purposes, that is, all the formulae would produce similar sized numbers, and small enough to illustrate graphically. For example, the curve extrapolates into very large numbers as compared to the trend because of the extra time coefficient. Therefore, the curve is restricted by less responsive coefficients than the trend. In fact, the slope starts off negative and then only increases by 0.1 multiplied by time squared. Random numbers ranging from 1 to 100 were added to these generated data, because this magnitude of random element is convenient, and sufficient to be graphically visible in the data over a span of 200 observations.

The random data themselves can be used to test the number of data required to recover the true average, 50.5. This is advantageous because truly random data should not have any underlying pattern which would mislead the recovery of the true average from a limited number of observations. Only if they are truly "random" can the retrieval of the average be successful. To ensure this, both a random number generation procedure, as well as a random number table from Fisher and Yates (1974: 134), were used as the sources for random data. Thus, two sets of data were generated for each of the four models: firstly, those which the Fisher data were added to, and secondly, those using the generated random numbers.

The procedure followed by Fisher and Yates in "generating" their random numbers, is shown in Appendix 5. Their method is, essentially, to take the 15-19th digits of numbers from a logarithm table, and then to make certain selections from those, "at random". They do not, however, explain how that random selection itself was made, "at random". The Fisher and Yates data will be referred to as the Fisher data from this point on.

Computer based random number generators, or pseudo random generators, used to create the other random series, make use of modulo division (remaindering method) to compute numbers. "...They are not real random numbers in the technical sense, since they are completely determined once the recurrence relation is defined and the parameters of the generator are specified..." (Winston, 1994: 1125). Such formulae, if properly executed, guarantee that each number occurs exactly once per cycle. In other words, all numbers are equally likely.

By contrast, *real* random numbers have two characteristics: not only does any number within the given range have as much chance of recurring as any other number, but also, the sequence of the numbers does not depend on any repetitive formula. An example would be rolling a die, or spinning an accurately weighted roulette wheel. Unfortunately, these physical procedures of generating random numbers are neither practical, nor do they produce numbers of the various magnitudes required in this context.

The mathematical process for generating random numbers is:

$$x_{i+1} = (ax_i + c) \text{ modulo } m \ (i = 0,1,2...)$$

x The initial value of x (x_0) is called the *seed*. The value of x_{i+1} equals the remainder after dividing $(ax_i + c)$ by m.

a The constant multiplier affects the sequence of data produced.

c The constant is used to change the mean of the result.

m This is the modulus constant, used to compute the remainder through division. The range of the output is limited from 1 to m-1.

i The number of iterations, that is, the number of random numbers to be generated.

(Source: Winston, 1994: 1124)

In short, the critical factors are the *modulator* (m), which controls the length of the series, and the *multiplier* (a) which controls what happens in the series. These factors determine the sequence of the data generated, and implicitly produce a given degree of randomness. For example, Table 8.3.1 shows random numbers generated between 1 and 4, that is, the modulator is equal to 5. The multiplier varies from 1 to 7.

Table 8.3.1: Random numbers with different multipliers (modulator 5)

		Iteration					
Multiplier	Seed	1	2	3	4	5	6
1	1	1	1	1	1	1	1...
2	(1	2	4	3)	1	2	4...
3	(1	3	4	2)	1	3	4...
4	1	4	1	4	1	4	1...
5	1	0	0	0	0	0	0...
6	1	1	1	1	1	1	1...
7	(1	2	4	3)	1	2	4...

(Source: Diegel, 1974: 2)

The random data generated repeat themselves after the third iteration. The first complete sequences are shown in parenthesis. It is visually obvious that the multipliers of 1 and 5 do not produce random sequences: 1 because a multiplier of 1 cannot change the sequence at all; 5 because when a multiplier is as large as the modulator, the remainder is inevitably zero. The multiplier of 6 has the same sequence as the multiplier of 1, and 7 is the same as 2. "...The last value in a complete cycle is the multiplier generating the same cycle in reverse..." (Diegel, 1974: 2). So, from the simple example in Table 8.3.1, a number of rules for random number generation can be established:

- the multiplier must be greater then 1 and less than the modulator;
- the modulator must be a prime number;
- a multiplier greater than the modulus generates the same sequence as the multiplier subtracted from the modulator;
- a given sequence repeats itself once the number of iterations exceeds the modulator.

(Source: Diegel, 1974: 2)

Specific multipliers and modulators can now be chosen in accordance with these rules. This choice is also dependent on the application in question. The advantage of generating random numbers is precisely that they can be made to suit the purpose. It was decided provisionally to work with 200 data: 100 forwards and 100 backwards. This would seem sufficient data to permit the accurate retrieval of the correct coefficients of even the most complicated polynomial model in this experiment, the curve. By repeating the series backwards one neutralises unwanted processes that may exist in the forwards part of the series.

To generate a series from 1 to 100, the modulator must be 101. The other critical factor determining the behaviour of the random series is the multiplier. The first requirement is that the multiplier "exhaust" the series, that is, it generates *all* numbers between one and one less than the modulator. For example, with a modulator of 5, multiplier 2 and 3 produced 1 2 4 3 and 1 3 4 2 respectively. All numbers are there, but with only four observations, it is difficult to say which series is more random. Most people would agree that 1 3 4 2 is more random than 1 2 4 3, the sequence 1 2 4 3 being quite regular. This is even clearer in examining various multipliers for a modulator of 101.

Table 8.3.2: Random numbers for multiplier 2, 11 and 18 (modulator 101)

2	4	8	16	32	64	27	54	7	14	28	56	11	22	44	88	75	49	98...
11	20	18	97	57	21	29	16	75	17	86	37	3	33	60	54	89	70	63...
18	21	75	37	60	70	48	56	99	65	59	52	27	82	62	5	90	4	72...

It is visually obvious that the multiplier of 2 produces a series with a pocket of regular progression at the outset, shown in red. Furthermore, other pockets of regular data for the multiplier of 2, are shown in blue and green. This can significantly affect the retrieval of the true coefficients, the random element having a regularity of its own. On the other hand, the random sequences produced by different multipliers, 11 and 18 for example, are such that it is not intrinsically obvious which is more random than another. This requires a more effective measure of randomness than simply looking at the series.

8.4 Using a moving period of one

In a *truly random* series, no single datum has any relationship to previous or subsequent ones. Therefore, because of this lack of pattern in the data, a moving period of 1 would be the worst possible response to such data. If the forecast approach is "my forecast for tomorrow is what happened today" (M=1), discussed in section 3.2, one would obtain the highest possible sigma. It would certainly be higher than the standard deviation of the *best* forecast for random data, that is, the mean. Thus in relating $\sigma\,(M=1)$ to $\sigma\,(all\ data)$, one obtains a measure of randomness. Expressed as a percentage difference, perfectly random data should have a factor of -100 percent. In this example, all suitable series should have an average of 50.5, that is, every datum between 1 and 100 is present exactly once.

Exactly half the 100 possible multipliers produce sequences that are repetitions of previous multipliers, only in reverse. These can be eliminated from consideration because the order of the sequence does not effect this application. It is visually obvious that the multipliers 1 and 100 are not random at all. After this simple process of elimination, one is left with 49 possible multipliers.

The sigma relation of these remaining multipliers with $M=1$ leaves 12 between the limits of -90 percent and -110 percent, that is, within ±10 percent of the worst possible relationship. There are numerous sets produced with a relation more negative than -100 percent, up to -300 percent for a multiplier of 100. However, multipliers falling short of the -100 percent criterion, are as useful as those exceeding it. Furthermore, the averages for these multipliers do not obtain the required mean of 50.5 and therefore, do not include all the numbers from 1 to 100, so they should be eliminated anyway. The multipliers which are validated by their $-\sigma$ relation and average are shown in Table 8.4.1.

Table 8.4.1: Results of $M=1$ relationship

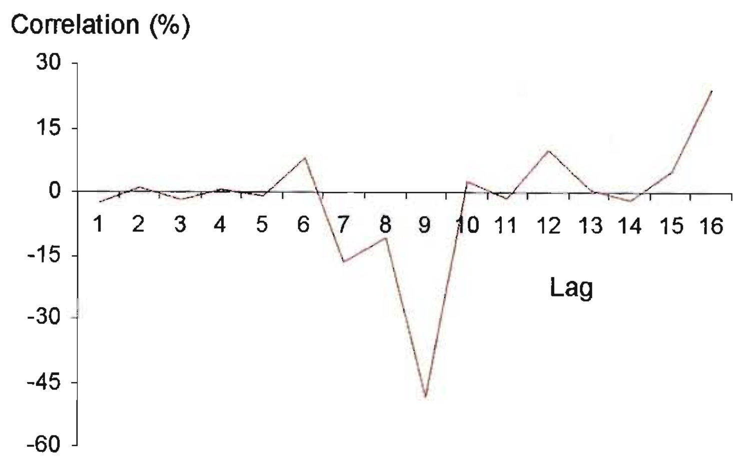
Multiplier	-σ Relation	Average
4	-106%	50.50
11	-95%	50.50
12	-90%	50.50
13	-108%	50.50
18	**-98%**	**50.50**
28	-105%	50.50
33	-106%	50.50
39	-109%	50.50
44	-91%	50.50
55	-106%	50.50
74	-109%	50.50

The multiplier of 18, show in bold, logically seems to be *best* because with -98 percent it is closest to the -100 percent mark . The relation of $M=1$ with the Fisher data was -58 percent with an average of 49.29. Because of the reputation of the authors in this field, their data will be used, although they do not fare particularly well in this randomness test.

8.5 Testing for periodicity

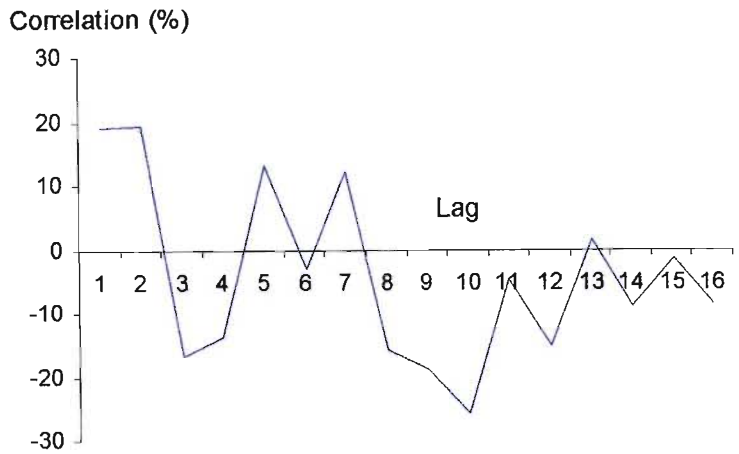
In addition to testing so-called random data with $M=1$, they can also be related to themselves, to determine that there is indeed no cyclical pattern in them. Autocorrelation was previously used in section 5.5 to determine the periodicity of a data series. Because random data should have no underlying process, they should have no periodicity either. The autocorrelation function for the random data with a multiplier of 18, can be seen in Figure 8.5.1.

Figure 8.5.1: Autocorrelation function for generated random numbers



There is a very slight periodicity of 2, indicated by the peaks with lag 2, 4, 6, 10 and 12. These peaks are so small, that is, the periodicity so slight, that the effect on the randomness of the series is negligible. The autocorrelation for the Fisher data is shown in Figure 8.5.2.

Figure 8.5.2: Autocorrelation function for Fisher data



Similarly with the Fisher data, there is practically no obvious periodicity. This provides some justification for its use as a random element in the construction of the generated data, after having fared relatively poorly with a moving period of one.

8.6 The transition matrix

In addition to the measures of randomness discussed, one may develop a new one by recalling the characteristics of an ideal random series. This is that events should be "random", unrelated to each other or independent of one another. As in rolling a die from trial to trial, one number should be as likely as any other. Numerically, this means that any given number should be followed by any other number, just as frequently as any other given number. In other words, one can actually count how often other numbers occur after any given number. More specifically, one can then see how often, for example, a low number such as 1 is followed by another low number, a higher one, or even a very high one. Indeed, one can study how the data *transit* from one point to the next.

The number of classes used to measure the transition of the data, for example, low, medium or high, needs to be informative. If there are too many groups, for example as many as there are data, the results say nothing of the transition distribution. On the other hand, if there are too few groups such as low, medium and high, this is not specific enough to be informative. Furthermore, the groups must be of equal size so that the transition results are not prejudiced towards large groups. The square root of the number of data gives a fair picture, although one less than this should be used so that there is a middle group in the case of 100 data. This suggests 9 groups of 11 data for the numbers ranging from 1 to 100.

The 100 data in Table 8.6.1 are those generated random numbers with a multiplier of 18, and will be used to construct a transition matrix, shown in Table 8.6.2.

Table 8.6.1: Generated raw data (Multiplier 18)

18	21	75	37	60	70	48	56	99	65	59	52	27	82	62	5	90	4	72	84
98	47	38	78	91	22	93	58	34	6	7	25	46	20	57	16	86	33	89	87
51	9	61	88	69	30	35	24	28	100	83	80	26	64	41	31	53	45	2	36
42	49	74	19	39	96	11	97	29	17	3	54	63	23	10	79	8	43	67	95
94	76	55	81	44	85	15	68	12	14	50	92	40	13	32	71	66	77	73	1

Table 8.6.2: Transition matrix for 100 generated data, 9 classes of 11

1	12	23	34	45	56	67	78	89	100	Sum
12	1	1	1	**2**	1	1	1	1	2	11
23	2	1	1	1	1	1	**2**	1	1	11
34	1	1	1	1	**2**	1	1	2	1	11
45	1	1	2	1	1	2	1	1	1	11
56	1	1	1	2	1	1	1	1	2	11
67	1	2	1	1	1	1	2	1	1	11
78	2	1	1	1	2	1	1	1	1	11
89	1	1	2	1	1	1	1	2	1	11
100	2	1	1	1	1	2	1	1	1	11
Sum	**12**	**10**	11	11	11	11	11	11	11	99

1385 sum squares off diagonal 3.7403 sigma for lag 1

The procedure for computing the transition matrix is mechanical and easily adaptable to a computer scenario. The first datum in the generated series is 18 (Table 8.6.1). The next is 21, which is classed in the same group as 18, group 1 of the possible 9, so 1 is added to the total for class 1 (cell [1,1]), the sum of numbers which remain in that group. The next datum is 75, classed in group 7 (67 to 78), so 1 is added to cell [1,7], and so the process continues. In a perfectly random series, all the cells in the matrix would have equal frequencies.

The transition matrix in Table 8.6.2 has a fairly even spread of frequencies for each class, most cells in the matrix being 1, and the sums of each row and column being almost a perfect 11. There are some two's and sums which are uneven (12 and 10), indicated in bold. This is due to the fact that there are only 9 groups for 100 data, meaning that some transitions will have to contain an extra entry.

Finally, a *sum of squares* and *sigma off-the-diagonal* can be computed to summarise the table. The sum-squares-off-diagonal is the figure which mathematically computes the evenness of the spread of transitions. Those transitions which fall on the diagonal from cell [1,1] to cell [9,9], shown in red in Table 8.6.2, represent transitions that remain in the same group, that is, they are not random but regular. To measure the "randomness" of transition, one looks at the evenness of distribution of transition *off* the diagonal.

The sum-squares-off-diagonal figure is the result of each transition multiplied by its squared position in the matrix. It is weighted by its distance from the diagonal, that is, by how irregular the data are. Specifically, the sigma is the sum-squares-off-the-diagonal divided by the number of transitions. Then the square root is taken to bring it back to the same order of magnitude as the transitions. This gives a clear picture of evenness, because it shows the average magnitude of the values weighted by their distance from the diagonal. The larger the sigma, the "wilder" the fluctuations in the data.

At the same time though, the spread of transitions should be even. A perfect transition, where all cells contain a 1, results in a sum-squares-off-the-diagonal of 1080 and a sigma of 3.6515. This means that, on the average, the data change between 3 and 4 groups with each transition. Remembering the 9 groups of 11, perfect data should change by a third with each transition. The generated data are not far off this mark, with 3.7403, substantiating their randomness.

Table 8.6.3 shows the first 100 of the Fisher raw data, also to be tested for evenness of transition.

Table 8.6.3: Fisher raw data

3	47	43	73	86	36	96	47	36	61	46	98	63	71	62	33	26	16	80	45
60	11	14	10	95	97	74	24	67	62	42	81	14	57	20	42	53	32	37	32
27	7	36	7	51	24	51	79	89	73	16	76	62	27	66	56	50	26	71	7
32	90	79	78	53	13	55	38	58	59	88	97	54	14	10	12	56	85	99	26
96	**96**	68	27	31	5	3	72	93	15	57	12	10	14	21	88	26	49	81	76

A simple inspection of the data in Table 8.6.3 reveals that there are numerous duplicate values. Checking these more carefully reveals that the numbers:

3, 12, 16, 24, 42, 47, 51, 53, 56, 57, 71, 73, 76, 79, 81, 88, 97

appear twice in the series. The values 7, 10, 27, 32, 36, 62 and 96 appear three times, with 96 appearing twice in a row, shown in bold in Table 8.6.3. The numbers 14 and 26 appear four times in the series. These data are, therefore, not exhaustive which explains why the average is not 50.5. The transition matrix for the Fisher data is shown in Table 8.6.4.

Table 8.6.4: Transition matrix for 100 Fisher data, 9 classes of 10.67

3.00	14	24	35	46	56	67	78	88	99	Sum
13.67	**3**	2	1	1	**4**	**0**	1	0	1	13
24.33	2	1	0	1	1	**3**	1	2	0	11
35.00	2	1	3	1	1	1	1	0	2	12
45.67	1	0	1	0	1	3	1	1	1	9
56.33	1	2	2	3	1	0	0	3	1	13
67.00	2	1	2	1	2	2	1	1	0	12
77.67	1	2	1	0	0	2	0	1	1	8
88.33	0	1	1	2	1	0	1	1	3	10
99.00	0	1	1	0	2	1	3	1	2	11
Sum	12	11	12	9	13	12	9	10	11	99

1060 sum squares off diagonal 3.2722 sigma for lag 1

In Table 8.6.4, the limits of the group sizes are slightly different from Table 8.6.2. This is because the numbers 1 and 2 do not occur in the series, and the first group therefore starts at 3. It is obvious, though, how different the spread of transitions is for these data. The size of the numbers in bold indicate that the Fisher data have more "pockets" of similar data than the generated data. This is highlighted by how many row- and column-sums differ from 11. The small sum-squares-off-the-diagonal and sigma further indicate this problem. Specifically, sigma 3.27 compares to 3.74 for the generated random numbers. This indicates that on the average, any new datum is more likely to be similar to the one just seen.

The transition matrices in Table 8.6.2 and 8.6.4 are computed for a lag of 1. This means that the transition of each datum is measured by comparing it to the next one in the series, *one* step away. This lag should be increased so that each datum is compared to the one 2 days hence, 3 days hence, and so on. These comparisons should also have an even frequency distribution. The comparison of different lags for the generated random data and Fisher can be seen in Table 8.6.5.

Table 8.6.5: Comparison of transition results with different lags

	Fisher		**Generated (Mult 18)**	
Lag	Sum Diag Sqr	Sigma	Sum Diag Sqr	Sigma
1	1060	**3.2722**	1385	3.7403
2	1247	3.5671	1268	3.5971
3	1552	4.0000	1350	3.7306
4	1596	4.0774	1264	3.6286
5	1388	3.8224	1317	3.7233
6	1368	3.8149	1215	3.5952
7	1079	3.4062	1432	3.9240
8	1267	3.7110	1275	3.7227
9	1240	3.6914	1851	**4.5101**
10	1160	3.5901	1198	3.6484
11	1285	3.7998	1208	3.6842
12	1403	3.9929	1092	3.5227
Total	15645	44.7453	15855	45.0272

The totals for each column show that the transition results for the two data series are of comparable magnitude. Some of the results, shown in bold in Table 8.6.5, are far from perfect, a measure of 4.51 indicating that data differ quite strongly over that interval, while 3.27 means that they are rather similar. Yet, on the whole, the results justify the use of these data as a "random" element.

The complete listing of all the generated data for the four models, can be seen in Appendix 6.

8.7 Summary

A comprehensive set of generated data are thus available for experimentation. These data include a random series of numbers between 1 and 100, which has been justified as "random" through objective testing. An identical set of data containing the Fisher random numbers is also available. It remains to discover the number of data it takes to retrieve the known coefficients of various models.

CHAPTER 9

Safety in numbers

9.1 Introduction

Using the generated data shown in Appendix 6, it will be attempted to determine the number of data required for accurate fitting of a given model. By fitting a limited number of observations and forecasting over a certain period, one can compute how successful the fit was. By increasing the number of data used in fitting, and again forecasting this new fit, the improvement can be related to the number of observations considered.

9.2 Retrieving the coefficients

To be even-handed in analysing the fitting process, two different random series were used for each model (those containing the generated random data, and then those with the Fisher data). Furthermore, the first 100 of the 200 generated data were divided into three overlapping groups, so as not to bias the investigation by starting the fitting by chance at a "pocket" of non-random data. The first group starts at datum 1, the second at datum 21 and the third at datum 41.

Three different size groups were used to fit a given model, then to compare the recovered coefficients to the true ones. "Small" groups will consider the first 15 data from the three starting points, "medium" groups will consider double that, the first 30 data from each point, and the "large" groups double again, 60 data. Thus each group has twice as many data as the previous one.

There are thus nine sets of individual data for each of the four models to be tested. The breakdown for all these sets is shown in Table 9.2.1.

Table 9.2.1: Group definition by starting point and size

Starting point	Small group	Medium group	Large group
1	1 - 15	1 - 30	1 - 60
21	21 - 35	21 - 50	21 - 80
41	41 - 55	41 - 70	41 - 100

A seven day time interval was used between each datum, in keeping with the client's practice of weekly data. Coefficients were computed by fitting all the data sets indicated in Table 9.2.1 for the generated data, and then testing them against the next 50 data, that is, roughly as if one was computing forecasts for the next year. The standard deviation was computed, and converted to a percentage of the lowest possible standard deviation, the one obtained when 100 data were used in fitting (the first complete cycle of the generated random numbers). This percentage indicates how closely the true coefficients were retrieved, 100 percent being the best possible result (with the retrieved coefficients identical to the original ones).

The results of fitting each model appear in the separate sections which follow. The results for *all* the data refer to the best fit, when using 100 data. The numbers 1, 21 and 41, in parenthesis, indicate the starting point of each of the three data groups, small, medium and large. Each entire series was tested both forwards and backwards. By testing the series in normal and reverse orders, one eliminates any regularity at the beginning or end of a series, which may accidentally improve or worsen the fit. However, those data which contain the computer generated random numbers have already been repeated backwards. Therefore, the fitting results for these data should have the same results when considering the series forwards or backwards.

9.2.1 Constant model

The average was the first model to be tested. The raw data for testing the average, that is, the generated random numbers from 1 to 100, with a multiplier of 18 and a modulator of 101, are shown in Table 9.2.1.1. The dividing line indicates the end of the first 100 data. Thereafter, data are repeated in reverse.

Table 9.2.1.1: 200 generated random data for fitting the average (forward and reverse)

18	21	75	37	60	70	48	56	99	65	59	52	27	82	62	5	90	4	72	84
98	47	38	78	91	22	93	58	34	6	7	25	46	20	57	16	86	33	89	87
51	9	61	88	69	30	35	24	28	100	83	80	26	64	41	31	53	45	2	36
42	49	74	19	39	96	11	97	29	17	3	54	63	23	10	79	8	43	67	95
94	76	55	81	44	85	15	68	12	14	50	92	40	13	32	71	66	77	73	1
1	73	77	66	71	32	13	40	92	50	14	12	68	15	85	44	81	55	76	94
95	67	43	8	79	10	23	63	54	3	17	29	97	11	96	39	19	74	49	42
36	2	45	53	31	41	64	26	80	83	100	28	24	35	30	69	88	61	9	51
87	89	33	86	16	57	20	46	25	7	6	34	58	93	22	91	78	38	47	98
84	72	4	90	5	62	82	27	52	59	65	99	56	48	70	60	37	75	21	18

Table 9.2.1.2 lists the Fisher random data, used as the second set of data in fitting the average. When 200 data are examined, one notes even more repetitions of certain observations:

1, 12, 16, 17, 22, 23, 30, 31, 35, 38, 39, 44, 54, 59, 63, 64, 67, 76, 77, 78, 82

appearing twice in addition to those listed in section 8.6, 96 and 30 appearing twice in successive observations (shown in bold in Table 9.2.1.2). The numbers 6, 43, and 55 appear three times.

Table 9.2.1.2: 200 Fisher data for fitting the average

3	47	43	73	86	36	96	47	36	61	46	98	63	71	62	33	26	16	80	45
60	11	14	10	95	97	74	24	67	62	42	81	14	57	20	42	53	32	37	32
27	7	36	7	51	24	51	79	89	73	16	76	62	27	66	56	50	26	71	7
32	90	79	78	53	13	55	38	58	59	88	97	54	14	10	12	56	85	99	26
96	**96**	68	27	31	5	3	72	93	15	57	12	10	14	21	88	26	49	81	76
55	59	56	35	64	38	54	82	46	22	31	62	43	9	90	6	18	44	32	53
23	83	1	**30**	**30**	16	22	77	94	39	49	54	43	54	82	17	37	93	23	78
87	35	20	96	43	84	26	34	91	64	84	42	17	53	31	57	24	55	6	88
77	4	74	47	67	21	76	33	50	25	83	92	12	6	76	63	1	63	78	59
16	95	55	67	19	98	10	50	71	75	12	86	73	58	7	44	39	52	38	79

The results of fitting the raw data from Tables 9.2.1.1 and 9.2.1.2, appear in Table 9.2.1.3.

Table 9.2.1.3: The results of fitting the average

Group (Start)	Average							
	Generated data				Fisher data			
	Forwards		Backwards		Forwards		Backwards	
	Sigma	%	Sigma	%	Sigma	%	Sigma	%
All	28.8661	100	28.8661	100	25.9319	100	30.0967	100
Small	15 data							
(1)	29.4791	103	29.4791	103	28.3193	109	28.7115	95
(21)	28.9914	101	28.9914	101	27.4295	106	27.4701	91
(41)	29.5951	103	29.5951	103	29.6549	114	25.9041	86
Average	29.3552 102%				27.9149 100%			
Medium	30 data							
(1)	29.7884	104	29.7884	104	26.9017	104	27.3895	91
(21)	28.8905	101	28.8905	101	30.3919	117	25.9213	86
(41)	29.7964	104	29.7964	104	29.6107	114	29.3091	97
Average	29.4918 103%				28.2540 101%			
Large	60 data							
(1)	29.6167	103	29.6167	103	28.6429	111	26.3927	88
(21)	29.7112	104	29.7112	104	29.1505	112	29.6788	99
(41)	28.8138	101	28.8138	101	25.9236	100	30.1483	100
Average	29.3806 103%				28.3228 101%			

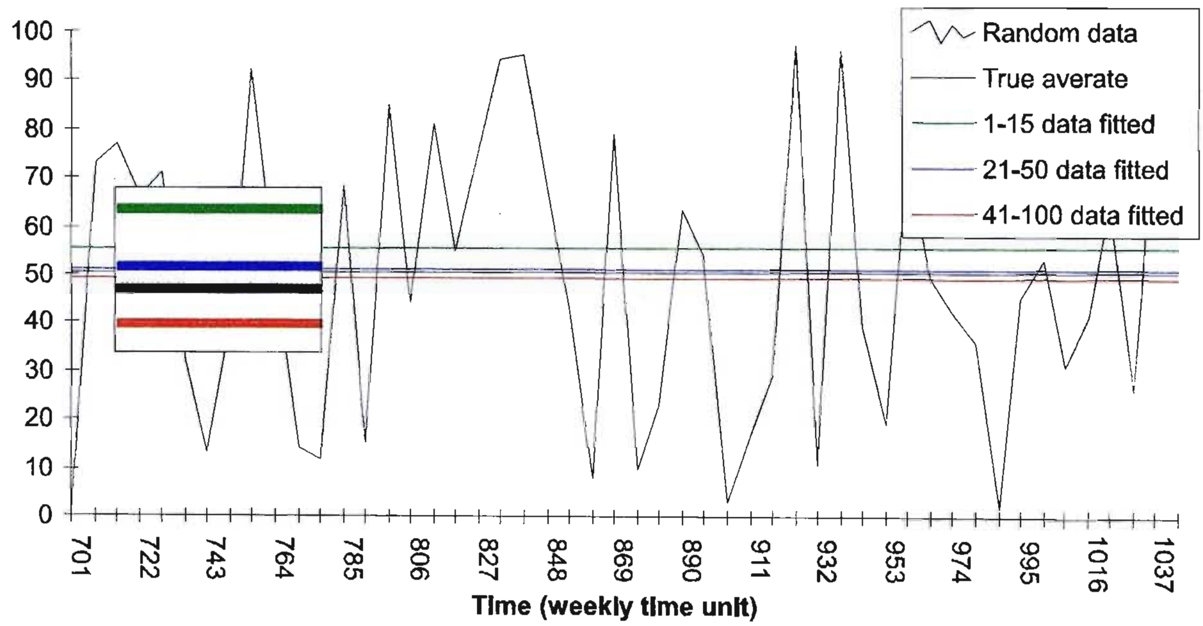
Obviously in Table 9.2.1.3, the average standard deviations for both data groups, generated and Fisher, do not improve as a larger group is used, that is, when the number of data used in fitting is doubled from 15 to 30, and then to 60 data. In other words, the results are remarkably similar no matter how many data were used in fitting. Both sets of data actually produced a slightly worse fit as more data are considered. This suggests that some parts of the series are not quite as "random" as others.

The fact that increasing the number of data does not improve fitting stands to reason: *random is random*, meaning that the data fluctuate from the mean by large amounts as often as they do by small amounts. Also, if the data *are* random, these fluctuations should be uniform throughout the series.

The mean should, therefore, be accurately retrieved from any part of the data, even from a small section, shaded in Table 9.2.1.3.

The same conclusion is corroborated by Figure 9.2.1.1, with the generated random numbers in black, and all the averages computed in colour. The green line indicates an average of a small group (1 to 15), the blue of a medium group (21 to 50), and the red of a large group (41 to 100). All the lines are close to the true average (shown in black), so much so that a zoom square is included in the graph.

Figure 9.2.1.1: Graph of averages with generated random numbers



9.2.2 Linear model

The next two series to be tested, are those for the trend. The generated data for the trend have an origin of 450 and slope of 10. The first series shown in Table 9.2.2.1 is that which contains the generated random numbers.

Table 9.2.2.1: Trend data with generated random element

478	491	555	527	560	580	568	586	639	615	619	622	607	672	662	615
710	634	712	734	758	717	718	768	791	732	813	788	774	756	767	795
826	810	857	826	906	863	929	937	911	879	941	978	969	940	955	954
968	1050	1043	1050	1006	1054	1041	1041	1073	1075	1042	1086	1102	1119	1154	1109
1139	1206	1131	1227	1169	1167	1163	1224	1243	1213	1210	1289	1228	1273	1307	1345
1354	1346	1335	1371	1344	1395	1335	1398	1352	1364	1410	1462	1420	1403	1432	1481
1486	1507	1513	1451	1461	1543	1557	1556	1571	1542	1533	1570	1632	1600	1574	1582
1648	1605	1685	1654	1701	1685	1716	1744	1755	1737	1723	1698	1779	1720	1743	1793
1794	1753	1777	1799	1877	1801	1896	1849	1839	1904	1889	1892	1896	1872	1925	1943
1931	1951	1984	1956	2020	2033	2060	1998	2004	2025	2030	2079	2108	2091	2049	2101
2147	2159	2113	2176	2116	2167	2140	2176	2165	2157	2166	2204	2238	2283	2222	2301
2298	2268	2287	2348	2344	2342	2284	2380	2305	2372	2402	2357	2392	2409	2425	2469
2436	2438	2470	2470	2457	2505	2461	2468								

Table 9.2.2.2 shows the second set of data for the trend containing the Fisher random numbers.

Table 9.2.2.2: Fisher data for the trend

463	517	523	563	586	546	616	577	576	611	606	668	643	661	662	643
646	646	720	695	720	681	694	700	795	807	794	754	807	812	802	851
794	847	820	852	873	862	877	882	887	877	916	897	951	934	971	1009
1029	1023	976	1046	1042	1017	1066	1066	1070	1056	1111	1057	1092	1160	1159	1168
1153	1123	1175	1168	1198	1209	1248	1267	1234	1204	1210	1222	1276	1315	1339	1276
1356	1366	1348	1317	1331	1315	1323	1402	1433	1365	1417	1382	1390	1404	1421	1498
1446	1479	1521	1526	1515	1529	1536	1525	1564	1548	1574	1612	1586	1572	1591	1632
1623	1599	1690	1616	1638	1674	1672	1703	1683	1753	1681	1720	1730	1726	1742	1807
1834	1789	1809	1824	1823	1844	1882	1827	1857	1923	1863	1928	1947	1905	1900	1986
1943	1994	1946	1964	2031	2014	2044	2012	1997	2043	2031	2067	2044	2085	2046	2138
2137	2074	2154	2137	2167	2131	2196	2163	2190	2175	2243	2262	2192	2196	2276	2273
2221	2293	2318	2309	2276	2365	2335	2357	2319	2408	2330	2380	2411	2425	2372	2456
2453	2448	2407	2454	2459	2482	2478	2529								

The procedure for testing the trend is exactly the same as with the average, and the results are shown in Table 9.2.2.3. The standard deviations when *all* the data were used in fitting, are practically identical to those in Table 9.2.1.3, as indeed the same random element was added to the trend line. The fact that they are not *absolutely* identical, shows that the random element itself has ever so slight a trend, as indeed is practically inevitable, even with a generated series.

The results for fitting the trend to the data containing the generated random element are identical whether they are fitted forwards or backwards, because of the arrangement of the random element, as previously discussed.

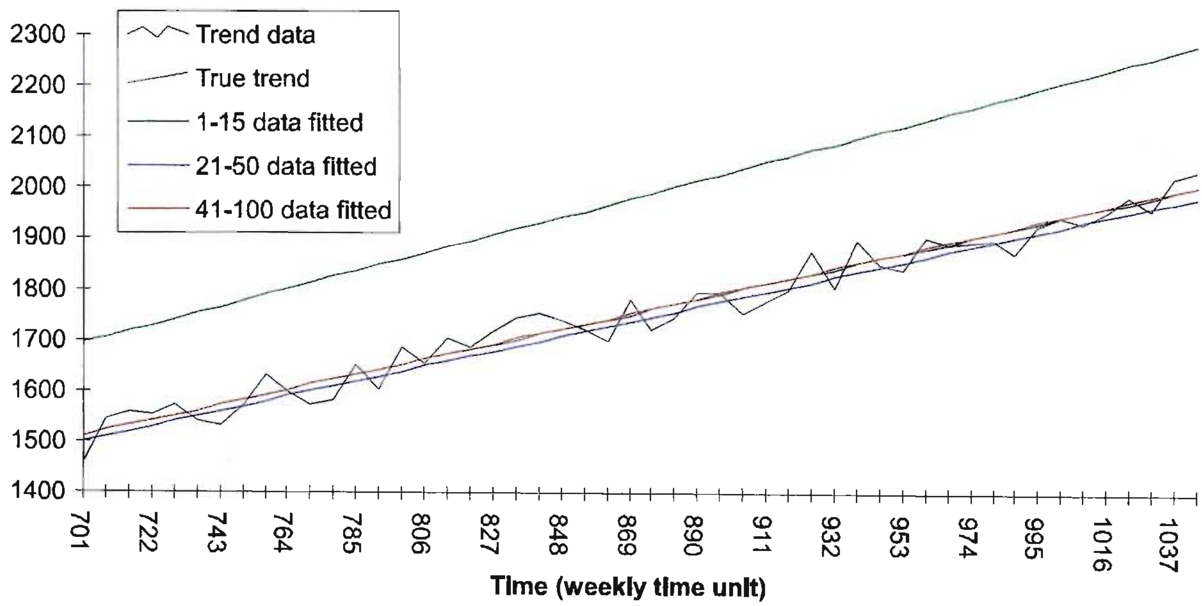
Table 9.2.2.3: The results of fitting the trend

Group (Start)	Trend (450 + 10 × time)							
	Generated data				Fisher data			
	Forwards		Backwards		Forwards		Backwards	
	Sigma	%	Sigma	%	Sigma	%	Sigma	%
All	28.8722	100	28.8722	100	25.9993	100	30.8608	100
Small	15 data							
(1)	79.7077	276	79.7077	276	91.7918	356	40.4810	131
(21)	**128.2926**	444	128.2926	444	27.1432	105	39.0697	127
(41)	42.5327	147	42.5327	147	**106.2492**	412	54.8460	178
Average	83.5110 289%				59.9302 211%			
Medium	30 data							
(1)	31.7902	110	31.7902	110	27.3590	105	28.6807	93
(21)	29.4547	102	29.4547	102	30.2144	116	27.0884	88
(41)	34.2674	119	34.2674	119	42.6125	164	33.8451	110
Average	31.8374 110%				31.6334 111%			
Large	60 data							
(1)	30.7450	106	30.7450	106	31.2682	120	26.8369	87
(21)	31.6857	110	31.6857	110	31.9658	123	29.9654	97
(41)	28.9729	100	28.9729	100	25.9715	100	31.5595	102
Average	30.4679 106%				29.5946 104%			

It is obvious from a cursory examination of Table 9.2.2.3, that the trend goes hopelessly wrong when only 15 data are considered. With so few data, the trend may *shoot off* on a slope wrongly retrieved from the initial period. In other words, when one deals with only about 15 so-called random numbers, they almost inevitably carry such a strong trend in themselves that they overpower the true trend. For example, the results 128.2926 and 106.2492 (shown in bold in Table 9.2.2.3) demonstrate just how wrong the forecast can be, if by chance, the fitting period and a non-random pocket coincide. Indeed, the second set of 15 data (21-35) in the generated random data, have an origin of 144.4 and a slope of -3.44, vastly different from the true origin of 450 and slope of 10. The generated random data are considerably worse with 15 data than the Fisher data, 83.5110 *versus* 59.9302 on average, because of this initial trendiness. It is exaggerated by the small fitting period.

However, there is an obvious improvement when 30 data are considered, shaded in Table 9.2.2.3. This means that the fitted slope is quite close to the true slope. This can be seen in Figure 9.2.2.1 which shows the slopes computed for 15 data (green), 30 data (blue), and 60 data (red), all having been fitted to the generated random numbers from point 1. The slopes are extended at their fitted rate of change for the next 50 weeks. The true slope of the data is shown in black, but is almost totally obscured by the fit in red which considers 60 data.

Figure 9.2.2.1: Graph of trends with generated random factor



There is a slight improvement in the slope retrieved when using 60 data rather than 30, producing an almost perfect fit. This improvement is so slight that it hardly justifies the extra amount of data, that is, twice as many. Although the trend is a more responsive model than the average, because of its rate of change, doubling the number of data from 15 to 30 counters practically all this responsiveness.

9.2.3 Curve model

The curve, however, is an even more responsive model than the trend because it has two rates of change. The raw data for the curve with the generated random data are shown in Table 9.2.3.1. The series is of a very similar magnitude to that shown in Table 9.2.2.1, the data used to test the trend.

Table 9.2.3.1: Curve data with generated random numbers

458	451	496	449	463	464	433	432	467	425	411	396	364	412	385	321
399	306	368	374	382	325	311	346	354	280	346	306	278	246	243	257
275	246	280	236	303	247	301	297	259	215	266	292	272	232	236	224
228	300	283	280	227	266	244	235	258	251	210	246	254	263	291	239
262	322	240	329	265	257	247	302	316	281	273	347	281	321	351	385
390	378	364	397	367	415	352	412	364	374	418	468	425	407	435	483
487	507	513	451	461	543	558	558	574	546	538	576	640	610	586	596
665	625	708	680	730	717	752	784	799	785	776	756	842	788	816	871
878	843	873	901	986	917	1019	979	976	1048	1041	1052	1064	1048	1110	1137
1134	1163	1205	1186	1260	1283	1320	1268	1285	1317	1333	1393	1433	1427	1397	1461
1519	1543	1510	1586	1539	1603	1589	1638	1641	1647	1670	1722	1771	1831	1785	1879
1891	1876	1911	1988	2000	2014	1973	2086	2028	2112	2159	2131	2184	2219	2253	2315
2301	2322	2373	2392	2398	2465	2441	2468								

Table 9.2.3.2 shows the data for the curve using the Fisher random data.

Table 9.2.3.2: Fisher data for the curve

443	477	464	485	489	430	481	423	404	421	398	442	400	401	385	349
335	318	376	335	344	289	287	278	358	355	327	272	311	302	278	313
243	283	243	262	270	246	249	242	235	213	241	211	254	226	252	279
289	273	216	276	263	229	269	260	255	232	279	217	244	304	296	298
276	239	284	270	294	299	332	345	307	272	273	280	329	363	383	316
392	398	377	343	354	335	340	416	445	375	425	388	395	408	424	500
447	479	521	526	515	529	537	527	567	552	579	618	594	582	603	646
640	619	713	642	667	706	708	743	727	801	734	778	793	794	815	885
918	879	905	926	932	960	1005	957	994	1067	1015	1088	1115	1081	1085	1180
1146	1206	1167	1194	1271	1264	1304	1282	1278	1335	1334	1381	1369	1421	1394	1498
1509	1458	1551	1547	1590	1567	1645	1625	1666	1665	1747	1780	1725	1744	1839	1851
1814	1901	1942	1949	1932	2037	2024	2063	2042	2148	2087	2154	2203	2235	2200	2302
2318	2332	2310	2376	2400	2442	2458	2529								

These results for fitting these data are shown in Table 9.2.3.3. In this case, the results for fitting the curve with the generated random element backwards and forwards are not identical. This is because the curve has a negative slope before it becomes positive, seriously misleading fitting in the forwards direction. Furthermore, the curve is comparatively far more responsive than the trend. Therefore, when the coefficients are not accurate, the fit goes even further off course. Only when a large number of data are used in fitting, is the forecast likely to have an acceptable degree of accuracy.

Table 9.2.3.3: The results of fitting the curve

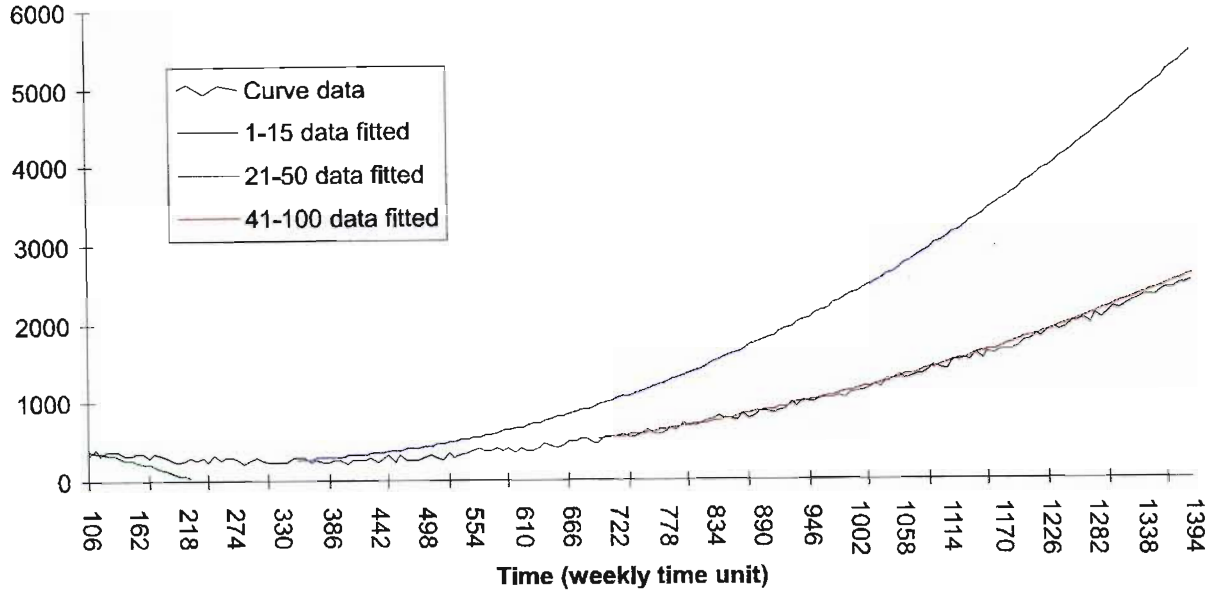
Group (Start)	Curve ($450 - 10 \times time + 0.1 \times time^2$)							
	Generated data				Fisher data			
	Forwards		Backwards		Forwards		Backwards	
	Sigma	%	Sigma	%	Sigma	%	Sigma	%
All	29.7077	100	29.6391	100	25.8049	100	31.8276	100
Small	15 data							
(1)	743.2700	2502	749.5997	2529	552.1503	2140	490.0752	1540
(21)	431.7519	1453	425.4487	1435	1158.3648	4489	139.8761	439
(41)	265.0554	892	271.3506	916	737.9029	2860	492.9411	1549
Average	481.0794 1621%				595.2184 2066%			
Medium	30 data							
(1)	247.7777	834	247.7500	836	**27.8872**	108	34.7388	109
(21)	221.1635	745	221.1981	746	178.2329	691	49.4260	155
(41)	64.5756	217	**64.5459**	218	181.4442	703	95.2081	299
Average	177.8351 599%				94.4895 328%			
Large	60 data							
(1)	50.7393	171	50.7412	171	29.7498	115	40.2998	127
(21)	34.0166	115	34.0317	115	39.8260	154	52.2140	164
(41)	32.5687	110	32.5091	110	65.6647	255	29.9729	94
Average	39.1011 132%				42.9545 149%			

Even with 30 data in the fitting period, the fit is very poor, because of the responsiveness of the model, and the changes occurring in the data. Within each group, the results vary considerably, except in the large group where the results are roughly of the same magnitude. For example, 64.5459 and 27.8872 (shown in bold) are comparatively good fits for only 30 data. This is because a slope is detected by chance in the fitting period that is not as poor as the same size group in another section of the data. Unfortunately, it is not possible to know beforehand which section of the series will closely reflect the true process.

The curve, being as sensitive as it is, means that any slope in the random data is amplified when forecasting a curve. The results of fitting a curve are largely dependent on whether or not these small regularities in the random data were encountered or not within the fitting period, and whether these regularities are similar to the true process or not. Because of its explosive nature, one has to be quite sure of the coefficients in the curve model before simply forecasting with it. When looking at 60 data, there is far greater consistency in the results, because the volume of data is sufficient to neutralise the random element and reveal the true process. Although there are still large errors when considering the largest group, shaded in Table 9.2.3.3, the reduction in comparison to the other groups is very large. It is obvious then that the more responsive the model, the more accurate the coefficients must be to keep the forecast on the right track. In order to retrieve the coefficients accurately, more and more data are required.

Three of the forecast curves are plotted in Figure 9.2.3.1, the raw data with the generated random element being shown in black. The fit from the first 15 data is shown in green, the first 30 data in blue, and the first 60 data in red. Each fit is cast forward for 50 weeks, to demonstrate its accuracy or inaccuracy. It is obvious how misleading the negative slope in first the 15 data was, the forecast reaching zero very quickly. It can also be seen that significantly more data are required to obtain a reasonably accurate estimate, than with the trend, due to the extra polynomial coefficient, time squared.

Figure 9.2.3.1: Graph of curves with generated random factor



9.2.4 Seasonal model

Apart from experimenting with polynomial models, trigonometric models are important where data are cyclical in tendency, as is to be expected with demand or sales data. Data with a 5-point cycle were tested to determine the number of data required to successfully retrieve the coefficients for this type of model. The raw data with the generated random element are shown in Table 9.2.4.1.

Table 9.2.4.1: Generated random numbers with 5-point wave

532	611	547	361	410	584	638	528	423	415	573	642	499	406	412	519
680	476	396	434	612	637	510	402	441	536	683	530	358	356	521	615
518	344	407	530	676	505	413	437	565	599	533	412	419	544	625	496
352	450	597	670	498	388	391	545	643	517	326	386	556	639	546	343
389	610	601	569	353	367	517	644	535	347	360	593	598	515	391	445
608	666	527	405	394	599	605	540	336	364	564	682	512	337	382	585
656	549	397	351	515	663	549	390	421	546	603	512	416	400	528	602
540	339	435	558	671	527	400	444	609	657	515	332	429	524	613	535
378	353	531	619	569	335	446	553	609	546	373	392	550	592	517	377
381	555	654	498	404	433	614	618	496	359	380	583	678	533	333	401
601	679	505	410	366	571	610	518	349	357	520	624	530	417	372	605
668	510	371	448	598	662	476	414	355	576	672	499	376	409	579	689
528	372	420	574	627	547	345	368								

The Fisher random data with a 5-point wave are shown in Table 9.2.4.2.

Table 9.2.4.2: Fisher random data with 5-point wave

517	637	515	397	436	550	686	519	360	411	560	688	535	395	412	547
616	488	404	395	574	601	486	334	445	611	664	496	391	412	556	671
486	381	370	556	643	504	361	382	541	597	508	331	401	538	641	551
413	423	530	666	534	351	416	570	640	498	395	357	546	680	551	402
403	527	645	510	382	409	602	687	526	338	360	526	646	557	423	376
610	686	540	351	381	519	593	544	417	365	571	602	482	338	371	602
616	521	405	426	569	649	528	359	414	552	644	554	370	372	545	652
515	333	440	520	608	516	356	403	537	673	473	354	380	530	612	549
418	389	563	644	515	378	432	531	627	565	347	428	601	625	492	420
393	598	616	506	415	414	598	632	489	377	381	571	614	527	330	438
591	594	546	371	417	535	666	505	374	375	597	682	484	330	426	577
591	535	402	409	530	685	527	391	369	612	600	522	395	425	526	676
545	382	357	558	629	524	362	429								

The results from testing the generated and Fisher data with a 5-point cycle, are shown in Table 9.2.4.3. The results for testing the series containing the generated random element backwards and forwards are different. This is understandable because 200 data are exactly divisible by the 5-point wave. This means that the series starts with a peak when fitted forwards, and a valley backwards. Therefore, one can expect the fitting results to be slightly different.

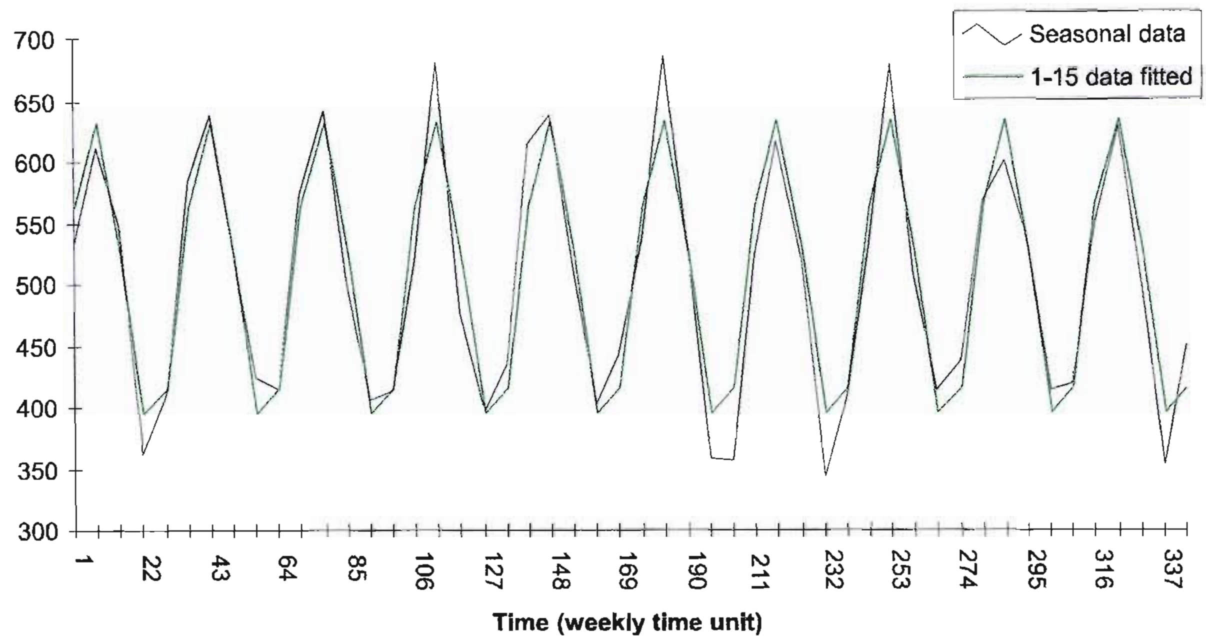
Table 9.2.4.3: The results of fitting a 5-point season

Group (Start)	Season: 5-point (450 + 100 sin (72°) × time - 100 cos (72°) × time)							
	Generated data				Fisher data			
	Forwards		Backwards		Forwards		Backwards	
	Sigma	%	Sigma	%	Sigma	%	Sigma	%
All	28.8685	100	28.8226	100	25.9186	100	30.5748	100
Small	15 data							
(1)	30.8179	107	30.7675	107	28.8496	111	28.9396	95
(21)	29.1015	101	29.1024	101	34.4080	133	28.2264	92
(41)	34.6528	120	34.6733	120	30.0995	116	29.2686	96
Average	31.5192 109%				29.9653 106%			
Medium	30 data							
(1)	29.7881	103	29.7685	103	29.6025	114	27.3831	90
(21)	31.9213	111	31.9628	111	32.4175	125	27.6438	90
(41)	30.1374	104	30.1363	105	29.9947	116	31.5331	103
Average	30.6191 106%				29.7625 105%			
Large	60 data							
(1)	30.8895	107	30.9250	107	29.6748	115	26.6889	87
(21)	29.8472	103	29.8483	104	28.6201	110	31.1337	102
(41)	30.2328	105	30.1890	105	26.4525	102	31.0151	101
Average	30.3220 105%				28.9309 102%			

The seasonal data are fitted well with only 15 data. This is because the only coefficients to be determined are those for the average, and for the periodicity. This agrees with Table 9.2.1.3, where 15 data have already been shown to be sufficient to retrieve the average. On the other hand, one can clearly discover the periodicity, and therefore recover the true sine and cosine coefficients, if the fitting period includes three full seasons, as shown in section 5.5. If there had been a trend with the season, one would expect to need at least 30 data to recover this polynomial slope, more than the 15 data required to determine the periodicity alone.

Figure 9.2.4.1 shows graphically how well the cyclical model coefficients were retrieved with only 15 data. The fit is shown in green, and the raw data containing the generated random factor in black.

Figure 9.2.4.1: Graph of 5-point season with generated random numbers



If the periodicity happened to be greater than 5, the periodicity and, therefore, the trigonometric coefficients could not be so successfully retrieved. Although one would expect the average to be retrieved from only 15 data, the sine and cosine coefficients for a periodicity larger than 5 will require more data. In particular, a lengthy cycle such as 12 weeks will require at least 36 data to determine the correct coefficients computationally. However, it is interesting to investigate the results when the periodicity is known, say 12, and forced on the fit using the different size groups from before. This example is shown in Table 9.2.4.4.

The results for fitting the data with generated random element forwards and backwards are identical, whereas they were different for the 5-point wave. With a periodicity of 12, the series starts and ends with a peak, which results in identical fitting backwards and forwards.

Table 9.2.4.4: The results of fitting a 12-point season

	Season: 12-point (450 + 100 sin (30°) × time - cos (30°) × time)							
	Generated data				Fisher data			
	Forwards		Backwards		Forwards		Backwards	
Group (Start)	Sigma	%	Sigma	%	Sigma	%	Sigma	%
All	28.8413	100	28.8413	100	26.2098	100	30.2051	100
Small	15 data							
(1)	32.3456	112	32.3456	112	29.0234	111	29.9068	99
(21)	34.8733	121	34.8733	121	32.4513	131	28.6192	95
(41)	30.8654	107	30.8654	107	33.8921	129	29.0132	96
Average	32.6948 113%				30.4843 108%			
Medium	30 data							
(1)	32.4172	112	32.4172	112	27.3742	104	28.6587	95
(21)	29.9444	104	29.9444	104	32.6700	125	26.2680	87
(41)	32.7109	113	32.7109	113	32.0968	123	29.9738	99
Average	31.6908 110%				29.5069 105%			
Large	60 data							
(1)	30.0746	104	30.0746	104	29.7354	114	26.4281	88
(21)	30.9907	107	30.9907	107	29.1196	111	29.6643	98
(41)	28.8126	100	28.8126	100	26.0708	100	30.1501	100
Average	29.9593 104%				28.5281 101%			

All the results are of very similar magnitudes. Certainly, when the periodicity is known, it only takes 15 data to establish the average, as shown before. On the other hand, if the data were to be fitted without the prior knowledge that they had a 12 week cycle, a fitting period of 60 data (a large group) would be required to retrieve the correct coefficients. This is because 12 multiplied by 3 is 36, larger than the data in a medium size group, but within the size of the large group. The critical factor for determining the number of data required in this case, is the period length, and not the order of the polynomial model, because the average can be recovered from even the smallest group.

The amount of data required to discover the true periodicity, and those required for fitting the polynomial part of the model, are independent and mutually exclusive. If there are insufficient data to discover the periodicity, the data cannot be accurately fitted. Similarly, 60 data are required for the curve, irrespective of the periodicity. If the fitting period does not consider enough data to satisfy both criteria, the result is a poor fit.

9.3 Conclusions from results

The results of the previous tables can be summarised in Table 9.3.1.

Table 9.3.1: Summary of best results for reasonable fitting periods

Model name	Polynomial Coefficients	Power of time	Period Length	Total Coefficients	Number of data required
Average	1	0	0	1	15
Trend	2	1	0	2	30
Curve	3	2	0	3	60
Short season	1	1	5	3	15
Long season	1	1	12	3	60

Some important conclusions can be drawn from the evidence in Table 9.3.1:

- as few as 15 observations are sufficient to establish the average;
- twice as many data are required to fit a trend rather than an average, and twice as many again to fit a curve rather than a trend;
- a season requires at least three times as many data as its periodicity.

Even with the presence of the random variation, the true average of a given set of data can be recovered from astonishingly few data. The results of fitting the generated data show that the number of data needs to be doubled for each extra polynomial term in the model. Also, three times as many data as the periodicity are required to fit a wave. The limitations when fitting a seasonal model can be either that the observations do not warrant the use of a high order polynomial model, or there are insufficient data to determine the optimal periodicity, or both.

If each new polynomial term requires double the number of data than for the previous model, then there is an exponential relationship between the two. Based on this relationship between observations and polynomial terms, Table 9.3.2 shows the extension for models with up to six polynomials, by doubling the amount of data required with each new term. The data required for the trigonometric model is similarly related to the number of polynomials contained in the model. However, just as important is the periodicity and whether it can be correctly discovered from the observations in hand.

Table 9.3.2: Data required for good fitting with polynomials

Polynomials	Highest power of time	Number of data
1	0	15
2	1	30
3	2	60
4	3	120
5	4	240
6	5	480

This projection suggests that data with a random variation comparable to that used in generating the test data, can be fitted accurately by having a number of observations as large as 15 times 2, raised to the power of the desired model. The power of the model is the highest exponent of *time*, for example, two for the curve (t^2). This formula is shown below.

$$\boxed{\text{Number of data} = 2^{\text{power of time}} \times 15}$$

Because the number of observations required by a higher order model doubles, the formula for the number of data required by a given model is *two* to the power of that model. From the empirical study in this chapter, 15 data is the size of the base group from which each subsequent group is doubled. Therefore, 2 to the power of the model multiplied by 15 will compute the number of data required to justify using a particular model. For business data, however, it is unlikely that one would want a model with a power of time higher than 2, because this would then be more adequately fitted by a logarithmic model. If one were to fit six polynomial coefficients, at least 480 data would be required.

Brown (1967: 119) has a different opinion on this point. "...If we use too short a historical period, the random fluctuations in demand may be misleading and yield coefficients that are not very good for forecasting. If we use too long a historical period, there is a cost of keypunching and processing a lot of data...As a rule of thumb many statisticians require that the number of observations be at least five times the number of terms in the model. Since we have to estimate six coefficients, that rule would suggest thirty observations...".

From the studies in this chapter, 30 observations is only enough to consider forecasting a trend, that is, two coefficients (t^0 and t^1). More importantly, the relationship between the required observations and the model is *exponential*, and not linear as suggested by Brown. With each higher order model, the number of observations must be doubled (15 30 60 120 240 480), not multiplied by the number of terms (5 10 15 20 25 30). As a rule, many more data than five times the number of terms is required, and specifically, one must also consider the *power* of the term.

9.4 Summary

The most important point is that an unexpectedly large number of observations are required to accurately fit a model other than the average. Even though high order models may fit the data well in the short term, extrapolating them into the future may have disastrous results. Obviously then, the more data that are available, the less the risk is of using a higher model mistakenly.

Any company hoping to accurately forecast a trend with their data, will have to keep nearly a year of weekly data or three years of monthly data. Fifteen observations does not seem like much, for example, just over two weeks of daily data. However, for those companies that store monthly or quarterly data, this can mean recording data for many years.

The analysis in this chapter ignored any revision of the forecast in the light of new data because this serves only to hide inaccurate model coefficients by keeping forecasts in line with the changing process. When forecasting into the true future though, for example, 50 weeks ahead, there are no data available to revise the forecast as they have not yet occurred. Also, the object of this study was to recover the *true* characteristics of a given process. Whether they are true can be seen, not by how well they fit the past, but how they forecast, unadulterated, into the future.

APPENDIX 1

Solving simultaneous equations by the substitution method

Equation 1: $A \times N + B \times \sum t = \sum D$

Equation 2: $A \times \sum t + B \times t^2 = \sum (t \times D)$

From Table 2.4.7, N=5, $\sum D$=150, $\sum(tD)$=540, $\sum t$=15, $\sum t^2$=55

Equation 1: $5A + 15B = 150$
Divide by 5 $A + 3B = 30$
$A = 30 - 3B$

Equation 2: $15A + 55B = 540$
Divide by 5 $3A + 11B = 108$

Substitute equation 1 into equation 2: $3(30\text{-}3B) + 11B = 100$

Multiply by 3: $90 - 9B + 11B = 108$
Solve for B: $2B = 18$
$B = \mathbf{9}$

Substitute solution for equation 2 into equation 1: $5A + 15(9) = 150$

Solve for A: $5A = 150 - 15(9)$
$A = \mathbf{3}$

APPENDIX 2

Listing of Pascal code for SHOWLOOP.PAS

```
program ShowLoop;
{ GD Armstrong, University of Natal, South Africa: July 1997 }

uses CRT;

const DataMatrix: array [1..3,1..5] of integer =  ((1, 1, 1, 1, 1),
                                                    (1, 2, 3, 4, 5),
                                                   (50,40,30,20,10));

{ DataMatrix is an array used to store Unity, Time and Data vectors,    }
{ which is only used to calculate the covariance matrix where data are }
{ stored in this fashion }

procedure DisplayDataMatrix;
{ Displays the data matrix which has been stored in memory }
var Line, Column: integer; { Variables to indicate position in DataMatrix }
begin
   ClrScr;
   WriteLn ('Data matrix stored in memory':30);
   for Line:= 1 to 3 do
   begin
      for Column:= 1 to 5 do Write (DataMatrix [Line, Column]:6);
      WriteLn
   end;
   WriteLn;
end;

procedure ComputeCovarianceWithDataStorage;
var CurrentLine, Line, Column: integer;
{ The current line in the data matrix is multiplied by all other lines }
    Hold: integer;
begin
   WriteLn('Covariance matrix: all data stored');
   for CurrentLine:= 1 to 3 do
   begin
      for Line:= 1 to 3 do
      begin
         Hold:= 0;
         for Column:= 1 to 5 do
          Hold:= Hold + DataMatrix [Line, Column]
                      * DataMatrix[CurrentLine, Column];
          { Multipy each line by the entire datamatrix }
         Write (Hold:6);
          { Output the results to the screen}
      end;
      WriteLn;
   end;
   WriteLn;
end;
```

```
procedure ComputeCovarianceWithoutDataStorage;
var Unity, Time, Data: integer;
{ Only the current values of the DataMatrix are stored for calculations }
    SumTime, SumUnit, SumData: integer;
{ The sums of these values also have to be stored }
begin
   Unity:= 1; { Unity vector is constant at one }
   Time:= 0; { Time starts a point zero }

   SumTime:= 0; SumUnit:= 0; SumData:= 0;
   WriteLn ('Covariance matrix: only current data storage');
   WriteLn ('N':6, 'Sum t**2':12, 'Sum D**2':12);

   repeat { Loop to compute covariance as data become available }
      Time:= Time + 1;
      { Increase time by one with each new datum }
      { Time is therefore generated and not stored }
      Read (Data); Gotoxy (Wherex, Wherey-1);
      { Input datum from user. Only one datum is stroed }
      SumUnit:= SumUnit + (Unity * Unity);
      SumTime:= SumTime + (Time * Time);
      SumData:= SumData + (Data * Data);
      { All the sums are computed with current Unit, Time and Datum }
      WriteLn (SumUnit:6, SumTime:12, SumData:12);
   until (Time = 5);
end;

begin
     DisplayDataMatrix;
     ComputeCovarianceWithDataStorage;
     ComputeCovarianceWithoutDataStorage;
     ReadLn;
end.
```

APPENDIX 3

Forecasting: data storage guidelines

1. The data best suited for accurate forecasting are daily data. This is because the number of working days per month, or per week, changes from one interval to another, and recording daily data overcomes this cause for inaccuracy. In any case, it is important that the *date* be recorded for each observation.

2. In having a date for each datum, missing data can easily be detected and dealt with. Data may be missing because of weekends, public holidays, strikes, stay-aways, and so on.

3. The format decided upon as most convenient and efficient for the time of each observation, is an integer date. For example, *19970101*, where the first four digits are the year (YYYY), the second two the month, (MM), and the final two, the date in that month (DD).

4. An observation is required for each stock keeping unit. Thus a six pack of *Product X* is recorded independently of the same *Product X* in a pack of three.

5. Demand data are better than sales data. Records from the point of sale provide a better picture of what is happening in the market than records of when new orders were placed.

6. The observations (demand) should appear after the date of the observation, and on the same line. This may also be an integer value.

7. Product identification could be provided by a heading/footing style message. Alternatively, this must be provided on each row, preceding the date, for example: BLUEPERS6 or BLUEPERS3. This facilitates a universally useful data base, but will waste space with the endless repetition of the product identification.

8. The name of each product is not required and a coded identifier may be substituted. This will improve the objectivity of the study by not knowing what the data are.

9. The data can be stored in a ASCII file. The end of a file should be marked with a slash "/", as some data bases do not write a proper EOF-signal.

10. Example of the ASCII file layout (the day of the week appears as a comment only):

```
Manufacturing Company - Product X: 6 pack
19980121          32          Wednesday
19980122          23          Thursday
19980123          40          Friday
19980126          38          Monday
/
```

It is clear and documented that the data for the weekend, Saturday (19980124) and Sunday (19980125), are missing.

APPENDIX 4

Restructured time unit

Time	Date	Time vector	Weeks	Gap
1	19920211	42	6	0
2	19920324	84	6	42
3	19920505	126	6	42
4	19920616	168	6	42
5	19920728	210	6	42
6	19920908	252	6	42
7	19921020	294	6	42
8	19921201	336	6	42
9	19921229	364	4	28
10	19930211	408	6	44
11	19930325	450	6	42
12	19930506	492	6	42
13	19930617	534	6	42
14	19930729	576	6	42
15	19930909	618	6	42
16	19931021	660	6	42
17	19931202	702	6	42
18	19931230	730	4	28
19	19940211	773	6	43
20	19940325	815	6	42
21	19940506	857	6	42
22	19940617	899	6	42
23	19940729	941	6	42
24	19940909	983	6	42
25	19941021	1025	6	42
26	19941202	1067	6	42
27	19941230	1095	4	28
28	19950204	1131	5	36
29	19950304	1159	4	28
30	19950401	1187	4	28
31	19950506	1222	5	35
32	19950603	1250	4	28
33	19950701	1278	4	28
34	19950805	1313	5	35
35	19950902	1341	4	28
36	19950930	1369	4	28
37	19951104	1404	5	35
38	19951202	1432	4	28
39	19951230	1460	4	28
40	19960204	1496	5	36
41	19960303	1524	4	28
42	19960331	1552	4	28
43	19960505	1587	5	35
44	19960602	1615	4	28
45	19960630	1643	4	28
46	19960804	1678	5	35
47	19960901	1706	4	28
48	19960929	1734	4	28
49	19961103	1769	5	35
50	19961201	1797	4	28
51	19961229	1825	4	28

APPENDIX 5

Computation of Fisher data

"...These random numbers are constructed from the 15th-19th digits of A.J. Thompson's 20-figure logarithm tables, parts IV, V, VI, VIII (40000-70000 and 80000-90000). The method of construction was as follows. Eight pages of 50 rows by 40 columns (afterwards reduced, for convenience in printing, to six pages of 50 rows by 50 columns) were each divided into 6 panels and 8 rows, there being thus 50 panels of 320 digits each in all (including two panels formed by the odd rows at the bottom of each page). To obtain each set of 50 digits a half page of logarithms (logarithms of 50 consecutive 5-figure numbers) was *selected at random*, and a column of digits (between the 15th and the 19th) was also *selected at random*, one digit being assigned to each of the 50 panels. The order of assignment was downwards on each page, but between each set of 50 digits the pages were rearranged *in a random order*..."

"...It will be seen that this method of construction can hardly fail to give a table of numbers which is the equivalent of a random selection made by an ideal mechanical contrivance. Any slight systematic element that may occur in parts of the logarithmic table will be effectively obliterated by the method of distribution of the digits..."

(Fisher and Yates, 1938: 18).

APPENDIX 6

Generated data

The *1* and *2* in the column headings denote the use of a generated random factor or the Fisher data respectively. The *Avg 1* numbers are the generated random numbers, modulator 101 and multiplier 18, and the *Avg 2* numbers are the Fisher data.

Date	Avg 1	Avg 2	Trend 1	Trend 2	Curve 1	Curve 2	5-Wave 1	5-Wave 2	12-Wave 1	12-Wave 2
19940101	18	3	478	463	458	443	532	517	431	416
19940108	21	47	491	517	451	477	611	637	508	534
19940115	75	43	555	523	496	464	547	515	625	593
19940122	37	73	527	563	449	485	361	397	624	660
19940129	60	86	560	586	463	489	410	436	647	673
19940205	70	36	580	546	464	430	584	550	620	586
19940212	48	96	568	616	433	481	638	686	535	583
19940219	56	47	586	577	432	423	528	519	469	460
19940226	99	36	639	576	467	404	423	360	449	386
19940305	65	61	615	611	425	421	415	411	378	374
19940312	59	46	619	606	411	398	573	560	372	359
19940319	52	98	622	668	396	442	642	688	402	448
19940326	27	63	607	643	364	400	499	535	440	476
19940402	82	71	672	661	412	401	406	395	569	558
19940409	62	62	662	662	385	385	412	412	612	612
19940416	5	33	615	643	321	349	519	547	592	620
19940423	90	26	710	646	399	335	680	616	677	613
19940430	4	16	634	646	306	318	476	488	554	566
19940507	72	80	712	720	368	376	396	404	559	567
19940514	84	45	734	695	374	335	434	395	497	458
19940521	98	60	758	720	382	344	612	574	448	410
19940528	47	11	717	681	325	289	637	601	360	324
19940604	38	14	718	694	311	287	510	486	351	327
19940611	78	10	768	700	346	278	402	334	428	360
19940618	91	95	791	795	354	358	441	445	504	508
19940625	22	97	732	807	280	355	536	611	509	584
19940702	93	74	813	794	346	327	683	664	643	624
19940709	58	24	788	754	306	272	530	496	645	611
19940716	34	67	774	807	278	311	358	391	621	654
19940723	6	62	756	812	246	302	356	412	556	612
19940730	7	42	767	802	243	278	521	556	494	529
19940806	25	81	795	851	257	313	615	671	438	494
19940813	46	14	826	794	275	243	518	486	396	364
19940820	20	57	810	847	246	283	344	381	333	370
19940827	57	20	857	820	280	243	407	370	370	333
19940903	16	42	826	852	236	262	530	556	366	392
19940910	86	53	906	873	303	270	676	643	499	466
19940917	33	32	863	862	247	246	505	504	520	519
19940924	89	37	929	877	301	249	413	361	639	587
19941001	87	32	937	882	297	242	437	382	674	619
19941008	51	27	911	887	259	235	565	541	638	614
19941015	9	7	879	877	215	213	599	597	559	557
19941022	61	36	941	916	266	241	533	508	548	523
19941029	88	7	978	897	292	211	412	331	501	420
19941105	69	51	969	951	272	254	419	401	419	401
19941112	30	24	940	934	232	226	544	538	343	337
19941119	35	51	955	971	236	252	625	641	348	364
19941126	24	79	954	1009	224	279	496	551	374	429
19941203	28	89	968	1029	228	289	352	413	441	502
19941210	100	73	1050	1023	300	273	450	423	587	560
19941217	83	16	1043	976	283	216	597	530	633	566
19941224	80	76	1050	1046	280	276	670	666	667	663
19941231	26	62	1006	1042	227	263	498	534	613	649
19950107	64	27	1054	1017	266	229	388	351	614	577
19950114	41	66	1041	1066	244	269	391	416	528	553
19950121	31	56	1041	1066	235	260	545	570	444	469
19950128	53	50	1073	1070	258	255	643	640	403	400
19950204	45	26	1075	1056	251	232	517	498	358	339
19950211	2	71	1042	1111	210	279	326	395	315	384
19950218	36	7	1086	1057	246	217	386	357	386	357
19950225	42	32	1102	1092	254	244	556	546	455	445
19950304	49	90	1119	1160	263	304	639	680	536	577
19950311	74	79	1154	1159	291	296	546	551	624	629
19950318	19	78	1109	1168	239	298	343	402	606	665
19950325	39	53	1139	1153	262	276	389	403	626	640
19950401	96	13	1206	1123	322	239	610	527	646	563

Date	Avg 1	Avg 2	Trend 1	Trend 2	Curve 1	Curve 2	5-Wave 1	5-Wave 2	12-Wave 1	12-Wave 2
19950408	11	55	1131	1175	240	284	601	645	498	542
19950415	97	38	1227	1168	329	270	569	510	510	451
19950422	29	58	1169	1198	265	294	353	382	379	408
19950429	17	59	1167	1209	257	299	367	409	330	372
19950506	3	88	1163	1248	247	332	517	602	316	401
19950513	54	97	1224	1267	302	345	644	687	404	447
19950520	63	54	1243	1234	316	307	535	526	476	467
19950527	23	14	1213	1204	281	272	347	338	510	501
19950603	10	10	1210	1210	273	273	360	360	560	560
19950610	79	12	1289	1222	347	280	593	526	666	599
19950617	8	56	1228	1276	281	329	598	646	595	643
19950624	43	85	1273	1315	321	363	515	557	593	635
19950701	67	99	1307	1339	351	383	391	423	554	586
19950708	95	26	1345	1276	385	316	445	376	508	439
19950715	94	96	1354	1356	390	392	608	610	444	446
19950722	76	96	1346	1366	378	398	666	686	389	409
19950729	55	68	1335	1348	364	377	527	540	368	381
19950805	81	27	1371	1317	397	343	405	351	431	377
19950812	44	31	1344	1331	367	354	394	381	457	444
19950819	85	5	1395	1315	415	335	599	519	572	492
19950826	15	3	1335	1323	352	340	605	593	565	553
19950902	68	72	1398	1402	412	416	540	544	655	659
19950909	12	93	1352	1433	364	445	336	417	599	680
19950916	14	15	1364	1365	374	375	364	365	564	565
19950923	50	57	1410	1417	418	425	564	571	537	544
19950930	92	12	1462	1382	468	388	682	602	505	425
19951007	40	10	1420	1390	425	395	512	482	390	360
19951014	13	14	1403	1404	407	408	337	338	326	327
19951021	32	21	1432	1421	435	424	382	371	345	334
19951028	71	88	1481	1498	483	500	585	602	421	438
19951104	66	26	1486	1446	487	447	656	616	479	439
19951111	77	49	1507	1479	507	479	549	521	564	536
19951118	73	81	1513	1521	513	521	397	405	623	631
19951125	1	76	1451	1526	451	526	351	426	588	663
19951202	1	55	1461	1515	461	515	515	569	588	642
19951209	73	59	1543	1529	543	529	663	649	623	609
19951216	77	56	1557	1536	558	537	549	528	564	543
19951223	66	35	1556	1525	558	527	390	359	479	448
19951230	71	64	1571	1564	574	567	421	414	421	414
19960106	32	38	1542	1548	546	552	546	552	345	351
19960113	13	54	1533	1574	538	579	603	644	326	367
19960120	40	82	1570	1612	576	618	512	554	390	432
19960127	92	46	1632	1586	640	594	416	370	505	459
19960203	50	22	1600	1572	610	582	400	372	537	509
19960210	14	31	1574	1591	586	603	528	545	564	581
19960217	12	62	1582	1632	596	646	602	652	599	649
19960224	68	43	1648	1623	665	640	540	515	655	630
19960302	15	9	1605	1599	625	619	339	333	565	559
19960309	85	90	1685	1690	708	713	435	440	572	577
19960316	44	6	1654	1616	680	642	558	520	457	419
19960323	81	18	1701	1638	730	667	671	608	431	368
19960330	55	44	1685	1674	717	706	527	516	368	357
19960406	76	32	1716	1672	752	708	400	356	389	345
19960413	94	53	1744	1703	784	743	444	403	444	403
19960420	95	23	1755	1683	799	727	609	537	508	436
19960427	67	83	1737	1753	785	801	657	673	554	570
19960504	43	1	1723	1681	776	734	515	473	593	551
19960511	8	30	1698	1720	756	778	332	354	595	617
19960518	79	30	1779	1730	842	793	429	380	666	617
19960525	10	16	1720	1726	788	794	524	530	560	566
19960601	23	22	1743	1742	816	815	613	612	510	509
19960608	63	77	1793	1807	871	885	535	549	476	490
19960615	54	94	1794	1834	878	918	378	418	404	444
19960622	3	39	1753	1789	843	879	353	389	316	352
19960629	17	49	1777	1809	873	905	531	563	330	362
19960706	29	54	1799	1824	901	926	619	644	379	404
19960713	97	43	1877	1823	986	932	569	515	510	456
19960720	11	54	1801	1844	917	960	335	378	498	541
19960727	96	82	1896	1882	1019	1005	446	432	646	632
19960803	39	17	1849	1827	979	957	553	531	626	604
19960810	19	37	1839	1857	976	994	609	627	606	624
19960817	74	93	1904	1923	1048	1067	546	565	624	643
19960824	49	23	1889	1863	1041	1015	373	347	536	510
19960831	42	78	1892	1928	1052	1088	392	428	455	491
19960907	36	87	1896	1947	1064	1115	550	601	386	437
19960914	2	35	1872	1905	1048	1081	592	625	315	348
19960921	45	20	1925	1900	1110	1085	517	492	358	333
19960928	53	96	1943	1986	1137	1180	377	420	403	446
19961005	31	43	1931	1943	1134	1146	381	393	444	456
19961012	41	84	1951	1994	1163	1206	555	598	528	571
19961019	64	26	1984	1946	1205	1167	654	616	614	576
19961026	26	34	1956	1964	1186	1194	498	506	613	621

Date	Avg 1	Avg 2	Trend 1	Trend 2	Curve 1	Curve 2	5-Wave 1	5-Wave 2	12-Wave 1	12-Wave 2
19961102	80	91	2020	2031	1260	1271	404	415	667	678
19961109	83	64	2033	2014	1283	1264	433	414	633	614
19961116	100	84	2060	2044	1320	1304	614	598	587	571
19961123	28	42	1998	2012	1268	1282	618	632	441	455
19961130	24	17	2004	1997	1285	1278	496	489	374	367
19961207	35	53	2025	2043	1317	1335	359	377	348	366
19961214	30	31	2030	2031	1333	1334	380	381	343	344
19961221	69	57	2079	2067	1393	1381	583	571	419	407
19961228	88	24	2108	2044	1433	1369	678	614	501	437
19970104	61	55	2091	2085	1427	1421	533	527	548	542
19970111	9	6	2049	2046	1397	1394	333	330	559	556
19970118	51	88	2101	2138	1461	1498	401	438	638	675
19970125	87	77	2147	2137	1519	1509	601	591	674	664
19970201	89	4	2159	2074	1543	1458	679	594	639	554
19970208	33	74	2113	2154	1510	1551	505	546	520	561
19970215	86	47	2176	2137	1586	1547	410	371	499	460
19970222	16	67	2116	2167	1539	1590	366	417	366	417
19970301	57	21	2167	2131	1603	1567	571	535	370	334
19970308	20	76	2140	2196	1589	1645	610	666	333	389
19970315	46	33	2176	2163	1638	1625	518	505	396	383
19970322	25	50	2165	2190	1641	1666	349	374	438	463
19970329	7	25	2157	2175	1647	1665	357	375	494	512
19970405	6	83	2166	2243	1670	1747	520	597	556	633
19970412	34	92	2204	2262	1722	1780	624	682	621	679
19970419	58	12	2238	2192	1771	1725	530	484	645	599
19970426	93	6	2283	2196	1831	1744	417	330	643	556
19970503	22	76	2222	2276	1785	1839	372	426	509	563
19970510	91	63	2301	2273	1879	1851	605	577	504	476
19970517	78	1	2298	2221	1891	1814	668	591	428	351
19970524	38	63	2268	2293	1876	1901	510	535	351	376
19970531	47	78	2287	2318	1911	1942	371	402	360	391
19970607	98	59	2348	2309	1988	1949	448	409	448	409
19970614	84	16	2344	2276	2000	1932	598	530	497	429
19970621	72	95	2342	2365	2014	2037	662	685	559	582
19970628	4	55	2284	2335	1973	2024	476	527	554	605
19970705	90	67	2380	2357	2086	2063	414	391	677	654
19970712	5	19	2305	2319	2028	2042	355	369	592	606
19970719	62	98	2372	2408	2112	2148	576	612	612	648
19970726	82	10	2402	2330	2159	2087	672	600	569	497
19970802	27	50	2357	2380	2131	2154	499	522	440	463
19970809	52	71	2392	2411	2184	2203	376	395	402	421
19970816	59	75	2409	2425	2219	2235	409	425	372	388
19970823	65	12	2425	2372	2253	2200	579	526	378	325
19970830	99	86	2469	2456	2315	2302	689	676	449	436
19970906	56	73	2436	2453	2301	2318	528	545	469	486
19970913	48	58	2438	2448	2322	2332	372	382	535	545
19970920	70	7	2470	2407	2373	2310	420	357	620	557
19970927	60	44	2470	2454	2392	2376	574	558	647	631
19971004	37	39	2457	2459	2398	2400	627	629	624	626
19971011	75	52	2505	2482	2465	2442	547	524	625	602
19971018	21	38	2461	2478	2441	2458	345	362	508	525
19971025	18	79	2468	2529	2468	2529	368	429	431	492

APPENDIX 7

Trend data with reduced random factor

The numbers in the column headings represent the multiplication factor used to reduce the relative size of the random factor in the simple trend with origin 450 and slope 10.

Date	Random	Trend 2	Trend 1	Trend 0.5	Trend 0.25
19940101	18	496	478	469	465
19940108	21	512	491	481	475
19940115	75	630	555	518	499
19940122	37	564	527	509	499
19940129	60	620	560	530	515
19940205	70	650	580	545	528
19940212	48	616	568	544	532
19940219	56	642	586	558	544
19940226	99	738	639	590	565
19940305	65	680	615	583	566
19940312	59	678	619	590	575
19940319	52	674	622	596	583
19940326	27	634	607	594	587
19940402	82	754	672	631	611
19940409	62	724	662	631	616
19940416	5	620	615	613	611
19940423	90	800	710	665	643
19940430	4	638	634	632	631
19940507	72	784	712	676	658
19940514	84	818	734	692	671
19940521	98	856	758	709	685
19940528	47	764	717	694	682
19940604	38	756	718	699	690
19940611	78	846	768	729	710
19940618	91	882	791	746	723
19940625	22	754	732	721	716
19940702	93	906	813	767	743
19940709	58	846	788	759	745
19940716	34	808	774	757	749
19940723	6	762	756	753	752
19940730	7	774	767	764	762
19940806	25	820	795	783	776
19940813	46	872	826	803	792
19940820	20	830	810	800	795
19940827	57	914	857	829	814
19940903	16	842	826	818	814
19940910	86	992	906	863	842
19940917	33	896	863	847	838
19940924	89	1018	929	885	862
19941001	87	1024	937	894	872
19941008	51	962	911	886	873
19941015	9	888	879	875	872
19941022	61	1002	941	911	895
19941029	88	1066	978	934	912
19941105	69	1038	969	935	917
19941112	30	970	940	925	918
19941119	35	990	955	938	929
19941126	24	978	954	942	936
19941203	28	996	968	954	947
19941210	100	1150	1050	1000	975
19941217	83	1126	1043	1002	981
19941224	80	1130	1050	1010	990
19941231	26	1032	1006	993	987
19950107	64	1118	1054	1022	1006
19950114	41	1082	1041	1021	1010
19950121	31	1072	1041	1026	1018
19950128	53	1126	1073	1047	1033
19950204	45	1120	1075	1053	1041
19950211	2	1044	1042	1041	1041
19950218	36	1122	1086	1068	1059

Date	Random	Trend 2	Trend 1	Trend 0.5	Trend 0.25
19950225	42	1144	1102	1081	1071
19950304	49	1168	1119	1095	1082
19950311	74	1228	1154	1117	1099
19950318	19	1128	1109	1100	1095
19950325	39	1178	1139	1120	1110
19950401	96	1302	1206	1158	1134
19950408	11	1142	1131	1126	1123
19950415	97	1324	1227	1179	1154
19950422	29	1198	1169	1155	1147
19950429	17	1184	1167	1159	1154
19950506	3	1166	1163	1162	1161
19950513	54	1278	1224	1197	1184
19950520	63	1306	1243	1212	1196
19950527	23	1236	1213	1202	1196
19950603	10	1220	1210	1205	1203
19950610	79	1368	1289	1250	1230
19950617	8	1236	1228	1224	1222
19950624	43	1316	1273	1252	1241
19950701	67	1374	1307	1274	1257
19950708	95	1440	1345	1298	1274
19950715	94	1448	1354	1307	1284
19950722	76	1422	1346	1308	1289
19950729	55	1390	1335	1308	1294
19950805	81	1452	1371	1331	1310
19950812	44	1388	1344	1322	1311
19950819	85	1480	1395	1353	1331
19950826	15	1350	1335	1328	1324
19950902	68	1466	1398	1364	1347
19950909	12	1364	1352	1346	1343
19950916	14	1378	1364	1357	1354
19950923	50	1460	1410	1385	1373
19950930	92	1554	1462	1416	1393
19951007	40	1460	1420	1400	1390
19951014	13	1416	1403	1397	1393
19951021	32	1464	1432	1416	1408
19951028	71	1552	1481	1446	1428
19951104	66	1552	1486	1453	1437
19951111	77	1584	1507	1469	1449
19951118	73	1586	1513	1477	1458
19951125	1	1452	1451	1451	1450
19951202	1	1462	1461	1461	1460
19951209	73	1616	1543	1507	1488
19951216	77	1634	1557	1519	1499
19951223	66	1622	1556	1523	1507
19951230	71	1642	1571	1536	1518
19960106	32	1574	1542	1526	1518
19960113	13	1546	1533	1527	1523
19960120	40	1610	1570	1550	1540
19960127	92	1724	1632	1586	1563
19960203	50	1650	1600	1575	1563
19960210	14	1588	1574	1567	1564
19960217	12	1594	1582	1576	1573
19960224	68	1716	1648	1614	1597
19960302	15	1620	1605	1598	1594
19960309	85	1770	1685	1643	1621
19960316	44	1698	1654	1632	1621
19960323	81	1782	1701	1661	1640
19960330	55	1740	1685	1658	1644
19960406	76	1792	1716	1678	1659
19960413	94	1838	1744	1697	1674
19960420	95	1850	1755	1708	1684
19960427	67	1804	1737	1704	1687
19960504	43	1766	1723	1702	1691
19960511	8	1706	1698	1694	1692
19960518	79	1858	1779	1740	1720
19960525	10	1730	1720	1715	1713
19960601	23	1766	1743	1732	1726
19960608	63	1856	1793	1762	1746
19960615	54	1848	1794	1767	1754
19960622	3	1756	1753	1752	1751
19960629	17	1794	1777	1769	1764
19960706	29	1828	1799	1785	1777
19960713	97	1974	1877	1829	1804
19960720	11	1812	1801	1796	1793

Date	Random	Trend 2	Trend 1	Trend 0.5	Trend 0.25
19960727	96	1992	1896	1848	1824
19960803	39	1888	1849	1830	1820
19960810	19	1858	1839	1830	1825
19960817	74	1978	1904	1867	1849
19960824	49	1938	1889	1865	1852
19960831	42	1934	1892	1871	1861
19960907	36	1932	1896	1878	1869
19960914	2	1874	1872	1871	1871
19960921	45	1970	1925	1903	1891
19960928	53	1996	1943	1917	1903
19961005	31	1962	1931	1916	1908
19961012	41	1992	1951	1931	1920
19961019	64	2048	1984	1952	1936
19961026	26	1982	1956	1943	1937
19961102	80	2100	2020	1980	1960
19961109	83	2116	2033	1992	1971
19961116	100	2160	2060	2010	1985
19961123	28	2026	1998	1984	1977
19961130	24	2028	2004	1992	1986
19961207	35	2060	2025	2008	1999
19961214	30	2060	2030	2015	2008
19961221	69	2148	2079	2045	2027
19961228	88	2196	2108	2064	2042
19970104	61	2152	2091	2061	2045
19970111	9	2058	2049	2045	2042
19970118	51	2152	2101	2076	2063
19970125	87	2234	2147	2104	2082
19970201	89	2248	2159	2115	2092
19970208	33	2146	2113	2097	2088
19970215	86	2262	2176	2133	2112
19970222	16	2132	2116	2108	2104
19970301	57	2224	2167	2139	2124
19970308	20	2160	2140	2130	2125
19970315	46	2222	2176	2153	2142
19970322	25	2190	2165	2153	2146
19970329	7	2164	2157	2154	2152
19970405	6	2172	2166	2163	2162
19970412	34	2238	2204	2187	2179
19970419	58	2296	2238	2209	2195
19970426	93	2376	2283	2237	2213
19970503	22	2244	2222	2211	2206
19970510	91	2392	2301	2256	2233
19970517	78	2376	2298	2259	2240
19970524	38	2306	2268	2249	2240
19970531	47	2334	2287	2264	2252
19970607	98	2446	2348	2299	2275
19970614	84	2428	2344	2302	2281
19970621	72	2414	2342	2306	2288
19970628	4	2288	2284	2282	2281
19970705	90	2470	2380	2335	2313
19970712	5	2310	2305	2303	2301
19970719	62	2434	2372	2341	2326
19970726	82	2484	2402	2361	2341
19970802	27	2384	2357	2344	2337
19970809	52	2444	2392	2366	2353
19970816	59	2468	2409	2380	2365
19970823	65	2490	2425	2393	2376
19970830	99	2568	2469	2420	2395
19970906	56	2492	2436	2408	2394
19970913	48	2486	2438	2414	2402
19970920	70	2540	2470	2435	2418
19970927	60	2530	2470	2440	2425
19971004	37	2494	2457	2439	2429
19971011	75	2580	2505	2468	2449
19971018	21	2482	2461	2451	2445
19971025	18	2486	2468	2459	2455

APPENDIX 8

Trend data with random element from different multipliers

The numbers in the headings for each trend (origin 450, slope 10), indicate the multiplier used to generate the random series, that is, 28, 11, 7 and 2.

Date	Trend 28	Trend 11	Trend 7	Trend 2
19940101	488	471	467	462
19940108	547	490	519	474
19940115	515	498	520	488
19940122	561	587	568	506
19940129	569	557	541	532
19940205	523	531	595	574
19940212	581	549	610	547
19940219	622	546	554	584
19940226	591	615	607	547
19940305	564	567	615	564
19940312	649	646	611	588
19940319	638	607	624	626
19940326	666	583	655	591
19940402	675	623	610	612
19940409	657	660	639	644
19940416	691	664	681	698
19940423	666	709	713	695
19940430	706	700	675	679
19940507	647	703	652	738
19940514	745	737	734	745
19940521	694	708	743	749
19940528	713	693	746	747
19940604	773	731	707	733
19940611	769	746	778	695
19940618	791	710	710	710
19940625	733	719	780	730
19940702	758	819	806	760
19940709	784	809	827	810
19940716	838	801	813	799
19940723	767	815	756	767
19940730	832	768	802	794
19940806	867	858	862	838
19940813	870	839	818	815
19940820	886	833	854	860
19940827	862	869	844	839
19940903	829	862	815	888
19940910	847	887	855	875
19940917	879	860	873	839
19940924	899	867	939	858
19941001	886	945	937	886
19941008	959	895	863	932
19941015	915	952	891	913
19941022	928	974	926	966
19941029	921	914	909	961
19941105	960	962	932	941
19941112	974	986	932	992
19941119	995	948	973	983
19941126	1010	935	998	955
19941203	958	995	1012	990
19941210	1050	1050	1050	1050
19941217	1033	1050	1054	1059
19941224	994	1051	1022	1067
19941231	1046	1063	1041	1073
19950107	1020	994	1013	1075
19950114	1032	1044	1060	1069
19950121	1098	1090	1026	1047
19950128	1060	1092	1031	1094
19950204	1039	1115	1107	1077
19950211	1090	1066	1074	1134
19950218	1137	1134	1086	1137

Date	Trend 28	Trend 11	Trend 7	Trend 2
19950225	1072	1075	1110	1133
19950304	1103	1134	1117	1115
19950311	1095	1178	1106	1170
19950318	1106	1158	1171	1169
19950325	1144	1141	1162	1157
19950401	1130	1157	1140	1123
19950408	1175	1132	1128	1146
19950415	1155	1161	1186	1182
19950422	1234	1178	1229	1143
19950429	1156	1164	1167	1156
19950506	1227	1213	1178	1172
19950513	1228	1248	1195	1194
19950520	1188	1230	1254	1228
19950527	1212	1235	1203	1286
19950603	1210	1291	1291	1291
19950610	1288	1302	1241	1291
19950617	1283	1222	1235	1281
19950624	1277	1252	1234	1251
19950701	1243	1280	1268	1282
19950708	1334	1286	1345	1334
19950715	1289	1353	1319	1327
19950722	1274	1283	1279	1303
19950729	1291	1322	1343	1346
19950805	1295	1348	1327	1321
19950812	1339	1332	1357	1362
19950819	1392	1359	1406	1333
19950826	1394	1354	1386	1366
19950902	1382	1401	1388	1422
19950909	1382	1414	1342	1423
19950916	1415	1356	1364	1415
19950923	1362	1426	1458	1389
19950930	1426	1389	1450	1428
19951007	1433	1387	1435	1395
19951014	1460	1467	1472	1420
19951021	1441	1439	1469	1460
19951028	1447	1435	1489	1429
19951104	1446	1493	1468	1458
19951111	1451	1526	1463	1506
19951118	1523	1486	1469	1491
19951125	1451	1451	1451	1451
19951202	1461	1461	1461	1461
19951209	1553	1516	1499	1521
19951216	1501	1576	1513	1556
19951223	1516	1563	1538	1528
19951230	1537	1525	1579	1519
19960106	1551	1549	1579	1570
19960113	1590	1597	1602	1550
19960120	1583	1537	1585	1545
19960127	1596	1559	1620	1598
19960203	1552	1616	1648	1579
19960210	1625	1566	1574	1625
19960217	1612	1644	1572	1653
19960224	1632	1651	1638	1672
19960302	1664	1624	1656	1636
19960309	1682	1649	1696	1623
19960316	1649	1642	1667	1672
19960323	1625	1678	1657	1651
19960330	1641	1672	1693	1696
19960406	1644	1653	1649	1673
19960413	1679	1743	1709	1717
19960420	1744	1696	1755	1744
19960427	1673	1710	1698	1712
19960504	1727	1702	1684	1701
19960511	1753	1692	1705	1751
19960518	1778	1792	1731	1781
19960525	1720	1801	1801	1801
19960601	1742	1765	1733	1816
19960608	1738	1780	1804	1778
19960615	1798	1818	1765	1764
19960622	1817	1803	1768	1762
19960629	1766	1774	1777	1766
19960706	1864	1808	1859	1773
19960713	1805	1811	1836	1832
19960720	1845	1802	1798	1816

Date	Trend 28	Trend 11	Trend 7	Trend 2
19960727	1820	1847	1830	1813
19960803	1854	1851	1872	1867
19960810	1836	1888	1901	1899
19960817	1845	1928	1856	1920
19960824	1873	1904	1887	1885
19960831	1862	1865	1900	1923
19960907	1947	1944	1896	1947
19960914	1920	1896	1904	1964
19960921	1889	1965	1957	1927
19960928	1930	1962	1901	1964
19961005	1988	1980	1916	1937
19961012	1942	1954	1970	1979
19961019	1950	1924	1943	2005
19961026	1996	2013	1991	2023
19961102	1964	2021	1992	2037
19961109	2023	2040	2044	2049
19961116	2060	2060	2060	2060
19961123	1988	2025	2042	2020
19961130	2060	1985	2048	2005
19961207	2065	2018	2043	2053
19961214	2064	2076	2022	2082
19961221	2070	2072	2042	2051
19961228	2051	2044	2039	2091
19970104	2078	2124	2076	2116
19970111	2085	2122	2061	2083
19970118	2149	2085	2053	2122
19970125	2096	2155	2147	2096
19970201	2129	2097	2169	2088
19970208	2129	2110	2123	2089
19970215	2117	2157	2125	2145
19970222	2119	2152	2105	2178
19970301	2172	2179	2154	2149
19970308	2216	2163	2184	2190
19970315	2220	2189	2168	2165
19970322	2237	2228	2232	2208
19970329	2222	2158	2192	2184
19970405	2177	2225	2166	2177
19970412	2268	2231	2243	2229
19970419	2234	2259	2277	2260
19970426	2228	2289	2276	2230
19970503	2223	2209	2270	2220
19970510	2301	2220	2220	2220
19970517	2299	2276	2308	2225
19970524	2323	2281	2257	2283
19970531	2283	2263	2316	2317
19970607	2284	2298	2333	2339
19970614	2355	2347	2344	2355
19970621	2277	2333	2282	2368
19970628	2356	2350	2325	2329
19970705	2336	2379	2383	2365
19970712	2381	2354	2371	2388
19970719	2367	2370	2349	2354
19970726	2405	2353	2340	2342
19970802	2416	2333	2405	2341
19970809	2408	2377	2394	2396
19970816	2439	2436	2401	2378
19970823	2374	2377	2425	2374
19970830	2421	2445	2437	2377
19970906	2472	2396	2404	2434
19970913	2451	2419	2480	2417
19970920	2413	2421	2485	2464
19970927	2479	2467	2451	2442
19971004	2491	2517	2498	2436
19971011	2465	2448	2470	2438
19971018	2517	2460	2489	2444
19971025	2478	2461	2457	2452

APPENDIX 9

Data supplied by the client

The product names have been removed at the request of the client. Each column heading represents a different stock keeping unit, although with a coded name. The *D* or *PP* in each column indicates whether the product is a Detergent or Personal Product. Each separate product has a different letter, and each derivative of the same product, still a separate stock keeping unit, has a different number.

Detergents

Date	DA1	DA2	DB1	DC1	DC2	DC3	DC4	DD1	DD2	DE1	DF1	DF2	DG1
19950204	94	82	122	334	217	393	27	91	65	72	193	185	37
19950304	194	100	124	338	205	333	21	163	87	74	167	192	34
19950401	89	80	139	396	195	388	23	118	73	77	171	216	48
19950506	87	74	113	399	223	450	30	103	81	83	174	240	38
19950603	103	58	112	327	190	335	30	116	53	74	146	195	26
19950701	71	68	95	415	229	463	29	91	51	88	204	184	7
19950805	115	79	125	625	346	621	40	128	79	104	167	257	7
19950902	100	64	119	379	211	416	23	76	61	81	147	231	2
19950930	147	75	154	279	136	332	36	122	61	82	189	294	71
19951104	131	80	158	574	229	611	46	127	77	118	182	263	52
19951202	92	63	114	332	242	361	53	109	58	102	152	151	34
19951230	77	70	102	374	212	393	23	88	54	92	141	176	38
19960204	81	73	109	389	249	421	44	84	49	122	160	191	34
19960303	101	90	82	480	257	484	38	50	55	105	177	207	31
19960331	118	83	91	410	273	431	43	139	75	101	136	203	34
19960505	163	108	110	437	234	461	34	171	96	134	180	193	40
19960602	64	61	93	359	181	400	28	91	55	120	169	328	40
19960630	42	68	89	444	194	493	22	56	63	117	232	272	41
19960804	109	83	102	622	354	656	47	119	61	95	169	195	46
19960901	90	74	104	404	194	386	30	118	50	123	140	160	37
19960929	110	60	102	440	133	437	28	101	52	94	142	157	41
19961103	151	97	212	691	293	746	49	161	86	145	258	336	63
19961201	111	82	75	402	243	392	40	139	51	84	124	156	32
19961229	89	72	82	290	162	331	36	100	59	94	133	137	42

Personal Products

Date	PPA1	PPA2	PPA3	PPA4	PPB1	PPB2	PPB3	PPC1	PPC2	PPD1
19920211	18946	4967	16615	38458	42360	44553	6899	18438	9741	875
19920324	30712	8719	27550	64981	83373	81231	11939	23122	15943	2183
19920505	21484	6399	23487	54427	67090	59083	8836	24652	14414	1190
19920616	19117	6805	12514	50719	87740	104504	13809	18285	12566	1533
19920728	18172	5638	20969	44837	103370	127806	17434	19727	9762	1439
19920908	15160	5553	17016	40540	81316	85970	12595	30098	11696	1174
19921020	17385	4746	10455	41896	49015	49949	11363	15721	9328	1939
19921201	22233	7608	18602	53320	48730	54077	9281	27781	14776	1176
19921229	15732	7082	23029	37040	34819	35220	6591	13392	8281	1258
19930211	22326	9484	28547	52198	41741	45278	7331	22640	12343	1181
19930325	35261	16294	38313	93069	91112	90051	12159	31375	13965	2003
19930506	41712	14202	36027	89741	53895	67321	10581	19691	9364	884
19930617	30168	14366	36669	77323	82043	90543	13060	18472	11676	5097
19930729	14312	7920	20979	30999	81580	83516	12402	17719	9754	4908
19930909	22028	11208	26259	48921	67656	77513	12343	21381	11252	4470
19931021	19864	9825	31965	49867	63507	64165	14032	19106	10211	6070
19931202	20214	10735	30209	48396	55794	70767	10623	28322	12266	5951
19931230	15176	8545	21712	35273	24986	31242	6100	7834	5734	4165
19940211	21069	7884	23707	52661	32209	42400	6538	16825	12705	6091
19940325	3741	5705	7109	61793	76499	69610	8950	23240	10382	6464
19940506	8681	3982	41247	71928	77399	89743	13723	16525	10329	4176
19940617	37646	21502	31555	77098	102326	107451	14800	22355	10794	1797
19940729	23582	11757	30682	70104	95501	100461	16355	22177	8750	6123
19940909	13067	6415	24117	35891	91127	95482	14935	17907	7360	6287
19941021	14316	7177	22236	35602	42858	51543	9898	20262	7316	4838
19941202	20768	13023	34377	54486	54651	64669	11499	22297	9368	5606
19941230	12799	5885	21898	22583	32920	33510	6797	16284	6103	4227
19950204	13904	4012	11503	27800	21490	28425	5708	8332	4256	3526
19950304	17497	6149	15497	39615	39177	41528	4484	13298	5682	5046
19950401	14081	7937	22932	34717	70072	66862	6271	21390	8260	4129
19950506	18169	7341	19696	46517	46537	50503	8123	13393	6518	4458
19950603	11944	6220	11877	22226	72491	67860	8544	13163	5614	3696
19950701	16163	5923	14859	35945	66103	68689	16677	14315	5033	3076
19950805	16396	7869	22588	40105	56800	64236	11223	13189	6700	4066
19950902	16684	5935	15031	44991	61160	65804	7075	10354	5290	3263
19950930	13953	5875	11524	29260	35658	32630	8105	16216	6154	3509
19951104	23456	8491	19093	50217	46429	53902	11117	19852	8417	5044
19951202	18739	9601	21968	42550	37794	52993	7103	16725	9491	4565
19951230	13139	6408	14781	28644	33763	32044	6509	11962	6635	3709
19960204	11774	4916	13447	29288	24038	29889	6120	11824	4224	3946
19960303	20335	7987	13259	43441	38059	38415	1886	884	1231	3699
19960331	17477	8853	20008	37786	58068	57062	1562	28	1200	4132
19960505	14945	7326	16977	36406	45919	58444	13224	28698	5511	4540
19960602	12338	5452	12362	17339	44162	60493	4709	8834	5896	2578
19960630	12830	5869	19522	27818	47803	61839	10452	16879	7259	4185
19960804	14482	6028	18201	51798	68171	74351	4727	22269	10021	4790
19960901	13583	4488	13965	48887	60189	68111	19	13518	8540	4149
19960929	12957	2542	9973	25923	39620	45484	3931	15508	7921	2998
19961103	17604	9916	20194	43572	77972	83449	15410	23121	11969	7079
19961201	13064	7633	20491	41945	23161	36829	8916	21387	11929	2687
19961229	14988	4389	13542	35171	19681	27574	5783	4507	8366	3784

BIBLIOGRAPHY

1. Brown, R.G. (1963) Smoothing, Forecasting and Prediction of Discrete Time Series. Englewood Cliffs: Prentice-Hall Inc.

2. Brown, R.G. (1967) Decision Rules for Inventory Management. New York: Arthur D. Little Inc.

3. Diegel, A. (1973) Iterative computation of exponential smoothing vectors. Paper presented at the conference of the American Institute of Decision Science, Boston, Massachusetts.

4. Diegel, A. (1974) A uniform random number generator with optimal characteristics. Paper presented at the conference of the Western branch of the American Institute of Decision Science, San Diego, California.

5. Diegel, A. (1987) From simultaneous equations to linear integer programming. Lecture notes, University of Natal, South Africa, Faculty of Commerce.

6. Fisher, R.A. & Yates, F. (1938) Statistical tables for biological, agricultural and medical research. London: Oliver and Boyd.

7. Fisher, R.A. & Yates, F. (1975) Statistical tables for biological, agricultural and medical research. London: Longman Group Ltd.

8. Kemeny, J.G., Schleifer, A., Snell, J.L. & Thompson, G.L. (1962) Finite Mathematics with Business Applications. Englewood Cliffs: Prentice-Hall Inc.

9. Kemeny, J.G., Schleifer, A., Snell, J.L. & Thompson, G.L. (1972) Finite Mathematics with Business Applications, Second edition. Englewood Cliffs: Prentice-Hall Inc.

10. Trueman, W. (1983) Applied time series and Box-Jenkins models. New York: Academic Press.

11. Winston, W.L. (1994) Operations research: Applications and Algorithms. Belmont, California: PWS-Kent.