

SPYROS MAKRIDAKIS

Forecasting accuracy and system complexity

Revue française d'automatique, d'informatique et de recherche opérationnelle. Recherche opérationnelle, tome 29, n° 3 (1995), p. 259-283.

http://www.numdam.org/item?id=RO_1995__29_3_259_0

© AFCET, 1995, tous droits réservés.

L'accès aux archives de la revue « Revue française d'automatique, d'informatique et de recherche opérationnelle. Recherche opérationnelle » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/legal.php>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

FORECASTING ACCURACY AND SYSTEM COMPLEXITY (*)

by Spyros MAKRIDAKIS ⁽¹⁾

Abstract. – Accurate forecasts are essential for a great number of applications yet large errors and considerable uncertainty characterize most of our attempts for predicting the future. This article surveys empirical studies on forecasting accuracy, summarizing their conclusions and examining the implications involved. It then discusses the effects of system complexity on forecasting accuracy and the related uncertainty while predicting future events. Finally, it provides suggestions for forecasting as accurately and realistically as possible when large and complex systems are involved.

Keywords: Forecasting accuracy, time series forecasting, empirical studies, system complexity, chaos theory.

Résumé. – Des prévisions précises sont indispensables dans un grand nombre d'applications et pourtant de grosses erreurs et une incertitude considérable caractérisent la plupart de nos tentatives de prédire l'avenir. Cet article examine des études empiriques sur la précision des prévisions en résumant leurs conclusions et en examinant leurs implications. Ensuite il traite de l'effet de la complexité des systèmes sur l'exactitude prévisionnelle et l'incertitude correspondante en prédisant les événements futurs. Pour conclure, il fournit des propositions pour prédire avec le plus de précision et de réalisme possible quand il s'agit de grands systèmes complexes.

Mots clés : Exactitude prévisionnelle, précision des prévisions, prévision de séries chronologiques, études empiriques, complexité des systèmes, théorie du chaos.

We might have trouble forecasting the temperature of coffee one minute in advance, but we should have little difficulty in forecasting it an hour ahead.

Edward Lorenz, *A Founder of Chaos Theory*

The 1960s were forecasters' euphoria. Computers, with exponentially increasing speed and decreasing costs, were promising unlimited and affordable power for solving practically any and all kinds of optimization problems. The belief at that time was of limitless potential, creativity being

(*) Received November 1993.

⁽¹⁾ Research Professor, INSEAD, boulevard de Constance, 77305 Fontainebleau Cedex, France.

the only constraint. Quantitative models, for example, growing bigger and in sophistication. This trend was nowhere as evident as in the areas of weather and economic forecasting, involving in many cases as many as 50000 equations. It was thought, at that time, that if small models can predict well, it is only natural to expect that bigger and more sophisticated ones would do even better.

Reality, however, did not turn out to satisfy these theoretical expectations. Bigger did not turn out to be better; more sophisticated did not improve accuracy. In the area of weather forecasting it soon became evident that no matter the size and sophistication of the models used, forecasting accuracy decreased considerably beyond two to three days and provided no better predictions than using the average weather conditions (referred to as climatology) of similar days of previous years to predict temperature, rainfall, or snow. What came to be known as the “butterfly effect” (sensitive dependence on initial conditions, such as those caused by the flying of a butterfly) could exert critical influence on future weather patterns which could not, therefore, be predicted since it was impossible to figure out the effect of the numerous initial conditions and how they could influence future weather conditions (Lorenz, 1966). In the short term too, the accuracy of weather forecasting could not improve much beyond the use of the naive approach which predicts that tomorrow’s or the next day’s weather will be exactly the same as today’s.

In economic and business forecasting, the accuracy of predictions did not turn out to be any better than those of weather forecasting. First, studies (McNees, 1975) comparing small and large econometric models showed that there was no difference between the two. Second, empirical comparisons (Armstrong, 1978) concluded that large and sophisticated econometric models were not more accurate than mechanical time series methods using a single equation. Finally, additional empirical work (Makridakis *et al.*, 1993; Fildes and Makridakis, 1995) indicated that simple methods did as well as or better than statistically sophisticated ones.

The purpose of this article is three-fold. First it surveys the literature of empirical studies related to the forecasting of economic and business series, outlining and examining their major conclusions. Second it discusses the effects of system complexity on forecasting accuracy and the related uncertainty. Finally, it provides suggestions for forecasting future events when large and complex systems are involved.

Model Fitting vs Post-Sample Forecasting Accuracy

In statistical forecasting, a model is fitted to the available data. The choice of the method (or methodology) is a matter of personal preference with some guidelines drawn from experience. Once a method (or methodology) has been selected, the specific model that *best fits* the available historical data for *one-period-ahead* forecasts is selected and used to predict the future (post-sample). This is realized by making at period t , m forecasts: X_{t+i} , $i = 1, 2, 3, \dots, m$. "Best fit" commonly means that the model minimizes the Mean Square Error (MSE), the Mean Absolute Percentage Error (MAPE), Mean Absolute Deviation (MAD), Median, Akaike's information criterion, or some analogous loss function. In some methods, such as ARIMA or Regression, the residual errors need to be independent, constant and normally distributed. In other methods (e. g., exponential smoothing, decomposition, Bayesian forecasting), no restrictions about these errors are made although it is desirable that they be random, constant and normally distributed. However, even if the residual errors resulting from fitting a model to available data are random, constant and normally distributed there is no way of guaranteeing that the post-sample forecasting errors will possess properties analogous to those of the model's disturbances. As a matter of fact empirical studies have shown that this is not ordinarily true (Makridakis and Winkler, 1989).

Two assumptions are implicit in the prevalent approach to model selection. First, it is assumed that the model that "best" fits the available data will also be the best model to predict beyond these data (post-sample). Second, it is assumed that the model that "best" forecasts for one period ahead, will also be best for predicting two, three, ..., m periods ahead. Both of these assumptions, however, do *not* hold true for many real-world economics/business series (Makridakis, 1990). The implications of the fact that the model that best fits the available data might *not* be the best model for post-sample forecasting have not been adequately considered in the forecasting literature (Priestley, 1979). Even during the 1970s the latter possibility was not mentioned in the most popular forecasting or econometric textbooks (Box and Jenkins, 1970; Montgomery and Johnson, 1976). Furthermore, no serious effort was made to validate the ability of the selected model to accurately forecast for out-of-sample periods. This is due partly because *all* data is being used to develop the "best" model, and partly due to the belief (originated in natural/physical sciences) that a "true best" model exists, and that such a model could be correctly identified and used for forecasting. In the social sciences most series used are short, measurement

errors abound, and controlled experimentation is not possible. Thus, the basic premise that the “best” model fitted to past data exists and can be identified, and is also best to forecast beyond these data, is invalid, and cannot be used as the basis for model selection. In other words the assumption of constancy of patterns and/or relationships (or even stationarity) does not hold in the majority of cases, for reasons to be discussed below.

If superior model fitting performance results in better out-of-sample forecasting accuracy then there should be a close correlation between the two. However, this is not the case; Makridakis (1986) and Makridakis and Winkler (1989) have found that such correlations are around 0.2 (*i.e.*, only 4% of the post-sample accuracy is explained) for the first three forecasting horizons, which then drop towards 0.1 by period five and zero by period 12. Similar conclusions (based on the M-Competition data) have been reached by Pant and Starbuck (1990). If a close relationship between model fit and out-of-sample forecasts does not exist it is hard to argue that model selection should be based on minimizing model fitting errors. Moreover, there is no reason to guarantee that a certain method will perform better than others because it better explains the past, or because its model fitting errors are smaller. In effect, the use of model fitting criteria may be viewed as more sophisticated variants of the case of fitting an $n - 1$ degree polynomial to a set of n data points to achieve zero model fitting errors. Forecasting errors will not be zero in practical applications.

THE CONSTANCY AND REGULARITY OF THE GENERATING PROCESS

Any system, no matter how simple or complex, can be presented by Figure 1 and consists of an input, a generating process and an output.

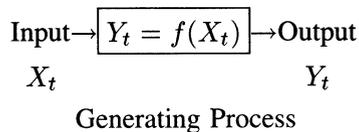


Figure 1. – The schematic presentation of input/output and generating processes

In forecasting terms the output Y_t needs to be predicted given the input X_t , knowing the generating process f , and realizing that the output can be influenced by fluctuations in both the input and the generating process.

By definition the impact of fluctuations, denoted by e_t , is considered to be white noise (otherwise it would be considered part of the input, or some systematic variation caused by the generating process itself which could be captured and included as part of the output) has a mean of zero, a constant variance, and is normally distributed,

$$E(e_t) = 0 \quad (1)$$

$$E(e_t e_{t+i}) = \begin{cases} \sigma^2 & \text{when } i = 0 \\ 0 & \text{when } i \neq 0 \end{cases} \quad (2)$$

Predicting Y_t can be realized by attempting to either relate the input to the output, which would require identifying and measuring their interconnections and the interrelationships among the various factors affecting the generating process; or alternatively it can be done by studying the actual behavior of Y_t over time. The first approach is called explanatory as it attempts to explain output through input and the generating process that transforms the latter to the former. Regression and econometric models are the major methods for applying this methodology. The second approach is called time series analysis. Its purpose is to identify and measure systematic patterns in Y_t which can consequently be extrapolated and used for forecasting.

In theory the generating process must be constant; otherwise the accuracy of forecasting could not be guaranteed. In practice, as it was mentioned above, this assumption of constancy has been ignored, or played down until rather recently (Fildes and Makridakis, 1995). The result has been that models have been developed with the only concern of how well they fitted available data, making measures such as R^2 or the minimization of the Sum of Square Errors (SSE) the only criterion for model selection. Consequently, constancy of relationships and/or patterns was assumed and equally accurate forecasts and uncertainty (expressed in terms of confidence intervals) were expected for future (post-sample) values not used when developing the model. In reality, however, relationships and patterns changed, sometimes considerably, during the post-sample phase. Such change makes forecasts and confidence intervals worse than expected through statistical theory (Makridakis *et al.*, 1987). This explains the common observation that simple models and the combining of forecasts of such models outperform, post-sample accuracy-wise, large and sophisticated ones.

Major finding of empirical studies

Comparative reviews of forecasting accuracy by alternative forecasting methods showed, as early as 1956 (Ferber (1956) and Shupack (1962)), that “correctly specified, well-fitted” models often underperform when tested outside the sample data. It was not until the late 1960s, however, that forecasting accuracy studies were performed on a wide range of comparisons that covered several methods and used real data in order to determine the post-sample accuracy of such methods. The greater availability, lower cost and improved speed and memory of computers made such empirical studies easier permitting an enlarged number of series and methods to be compared. Reid (1969), and Newbold and Granger (1974) using 106 time series concentrated on the accuracy of univariate forecasting methods. They also considered in depth various forms of combining forecasts from these three methods. The Makridakis and Hibon (1979) study was based on 111 series. While the Newbold and Granger study offered findings that confirmed (albeit weakly) the strongly held opinions of the leading time series statisticians of the day, the results of the Makridakis and Hibon study were more openly in conflict with such conclusions. Further study by Makridakis and his colleagues resulted in the so called M-Competition (Makridakis *et al.*, 1982) which reconfirmed the results of the Makridakis and Hibon study. Since 1982 many additional empirical comparisons (Fildes and Makridakis, 1995) have been conducted and the following conclusions reached.

Sophisticated vs. simple methods

Because sophisticated methods permit the user to select an appropriate theoretical model to best match the characteristics of the data, it could be reasonably expected that the theoretical correctness and the additional flexibility would lead to improved accuracy. However, this has not happened. Simple, mechanical methods, such as exponential smoothing or even the benchmark methods of a random walk, or a seasonal variant, outperformed, on average, more complex specifications such as the ARIMA methodology or variants of Kalman filters. Additional empirical research has given further support to the view that simple time series models do at least as well as statistically sophisticated ones: Schnaars (1986), Koehler and Murphee (1988), Huss (1985), Geurts and Kelly (1986), Watson *et al.* (1987), Collopy and Armstrong (1992), Fildes (1983), Makridakis *et al.* (1993). These studies use different data sets and a variety of methods.

General and special cases

It can be shown that many variants of exponential smoothing are special cases of ARIMA models (Cogger, 1974, Gardner and McKenzie, 1985). It was therefore argued that it was not possible for the latter to outperform the former, performance should at best be equal. Starting with Makridakis and Hibon (1979) evidence has accumulated that ARIMA based forecasting, even when applied strictly according to the authors' original intentions, is outperformed by exponential smoothing. This result was supported by Makridakis *et al.* (1982) and (1993). However, the only claim that can be made is that the model for fitting errors in ARIMA models will be as small as those of exponential smoothing methods which empirical studies have shown to be true. As far as post-sample accuracy is concerned, we cannot assure that more general models will be at least as accurate as their special case ones when the assumption of constancy does not hold true.

Forecasting accuracy and prediction intervals

All forecasts are expected of the most likely outcome and are associated with an interval that expresses uncertainty. Empirical studies (Makridakis and Winkler, 1989, Makridakis *et al.*, 1987) have shown however, that actual forecasts fall outside the theoretically constructed confidence intervals more often than postulated by the theory, regardless of the model used to describe the data (*see* Chatfield, 1993, for a survey of this topic). For example, in Makridakis *et al.* (1987) it was shown that 17% of the forecasts fell outside the 95% confidence interval for lead 1 rising to 26% for lead 6. This finding again means that constancy between model fitting and post-sample forecasts does not hold.

The effect of sample size

According to statistical theory, the size of the prediction interval, directly related to the standard error, ought to decrease according to the square root of the sample size used in model estimation. Empirically, however, this result has not been confirmed (Makridakis and Hibon, 1979; Makridakis *et al.*, 1982; and Lusk and Neves, 1984). Such a result contradicts a major tenet of statistical theory and raises some fundamental questions about the optimal sample size in real-life applications, the obvious conclusion being that there is no need to search for more data (an expensive and time consuming task) when less can do as well.

Combining

Starting with Newbold and Granger (1974) and supported by Makridakis and Hibon (1979), the M-Competition (1982), the M2-Competition (1993) and a large number of other empirical studies both inside and outside the field of forecasting (*see* Clemen, 1989, for a review and annotated bibliography) the conclusion has been maintained that combining more than one forecasting method (at least one of which will naturally be suboptimal for a given data set) results in more accurate out-of-sample forecasts. In addition, to add “insult to injury”, empirical research has found that simply averaging the forecasts of the various methods is as accurate as combining them according to some optimizing procedures that minimize the variance and/or covariance of the methods being combined.

THE “REAL” GENERATING PROCESS

In theory it is convenient to assume that the generating process is constant. In reality it is practically impossible to assure constancy. Figure 2 shows the average yearly temperatures in Paris since 1851, the earliest time for which data is available. Figure 2(a) shows a rising trend of about 0.016 degrees a year, which can be shown to be statistically greater than zero for the entire period. Alternatively, it can also be demonstrated statistically (*see* Figure 2(b)) that there is a yearly increase of 0.005 degrees between 1851 and 1920, and an even greater increase of 0.028 degrees from 1921 onwards.

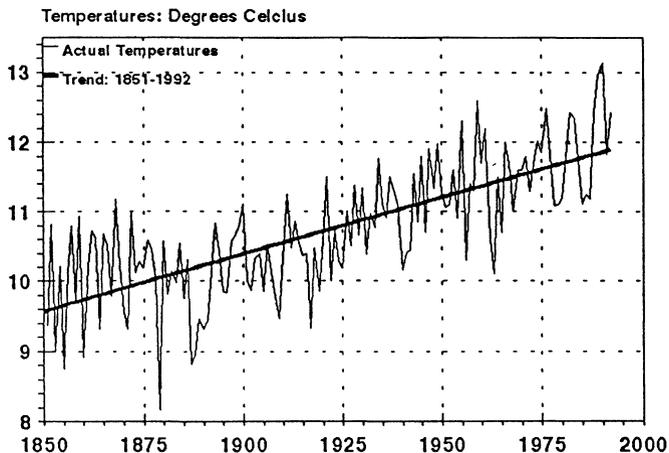


Figure 2(a). – Average yearly Paris temperatures – A single rising trend.

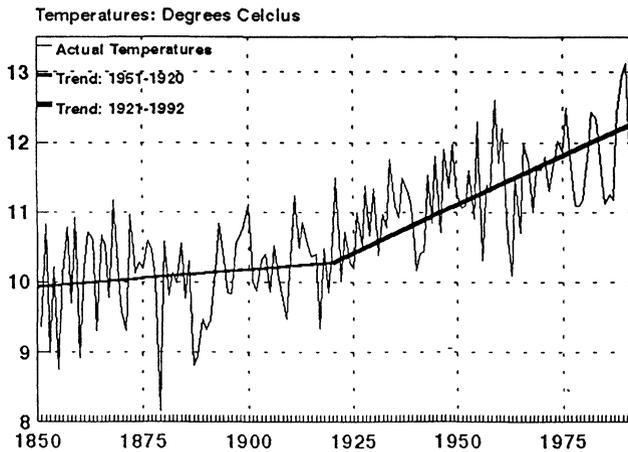


Figure 2(b). – Average yearly Paris temperatures – Two different rising trends.

Does Figure 2 imply a change in the generating process (that is, the Earth's climate) that will result in global warming? Answering this question is not possible without additional information as a temperature rise for 142 years can be the increasing part of a long cycle, or be the result of temporary factors (e.g., the urbanization of Paris).

For forecasting purposes our predictions will be quite different if we assume that:

(a) the generating process has changed, as suggested in Figure 2, implying continued increase in the Earth's temperature and a global warming

(b) the rise in temperature seen in Figure 2 is part of a cycle that will peak and reverse itself

(c) urbanization has contributed to the temperature increase seen in Figure 2 and as such urbanization has been completed and effective pollution control is reducing heat exhausts, Paris temperatures will stop their increase.

The forecasting of business and economic series can also provide quite different predictions depending upon our assumptions concerning the behavior and constancy of the generating process involved.

Figure 3 shows real, that is, adjusted for inflation, copper prices for four different periods, each longer than the previous one. The obvious conclusion from Figure 3(a) is that prices are falling, from Figure 3(b) that they are constant, from Figure 3(c) that they are increasing, while from Figure 3(d) that they are declining. As each graph of Figure 3 covers a certain time span, we have to be sure that the generating process that applies to such a span is constant.

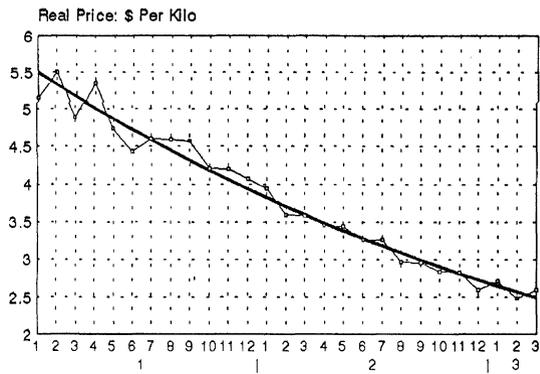


Figure 3(a). - Real 1994 monthly copper prices.

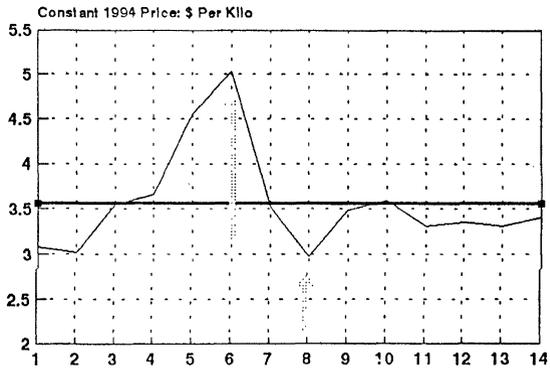


Figure 3(b). - Real yearly copper prices.

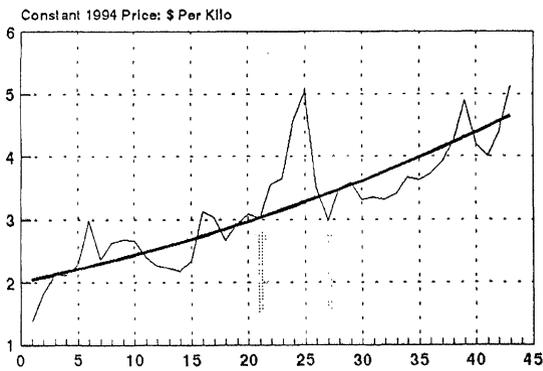


Figure 3(c). - Real yearly copper prices.

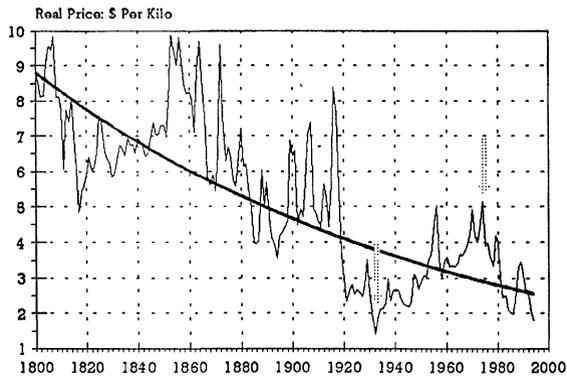


Figure 3(d). – Real copper prices in constant 1994 \$.

Clearly, the longer the time horizon the stronger the chance of describing the real generating process and not showing some cyclical fluctuation of a certain length and amplitude, as cycles are widespread in all types of phenomena from natural and physical to economic and business ones. This means that the real generating process cannot be inferred unless we can persuade ourselves that any persistent increase or decrease is not part of some cycle. This is obviously the case in Figures 3(a), 3(b) and 3(c) which describe only a small, cyclical part of the data since 1800 displayed in Figure 3(d). A further study of the copper price data can assure us that 1800 is the right time to start (this is when the effects of the industrial revolution started being felt, see also below). In such a case we can conclude that the exponential decreasing prices and the huge cycles of various lengths and amplitudes are part of the real generating process and must be included in any and all forms of forecasting beyond 1993. In other words, and until evidence proving otherwise becomes available, we can conclude that Figure 3(d) depicts the real generating process of real copper prices.

For the Paris temperatures we cannot make a similar claim because temperature cycles lasting several thousand years have existed throughout the Earth's history, bringing both ice ages and hot summers to places covered by ice beforehand. Thus, we cannot infer the real generating process of temperatures for the Paris region unless much more data is available to us that needs to be found through other sources such as ice core, pollen, and deep ocean measurements (Duplessy and Morel, 1990) which suggest cyclical fluctuations considerably bigger than those in the Paris temperatures over the last 143 years (Jouzel *et al.*, 1987, Guiot, 1989).

Random walks and cyclical movements

Figures 4(a), 4(b), 4(c) and 4(d) show four series generated as follows:

$$Y_t = \sum e_t \quad (3)$$

where e_t is defined by (2).

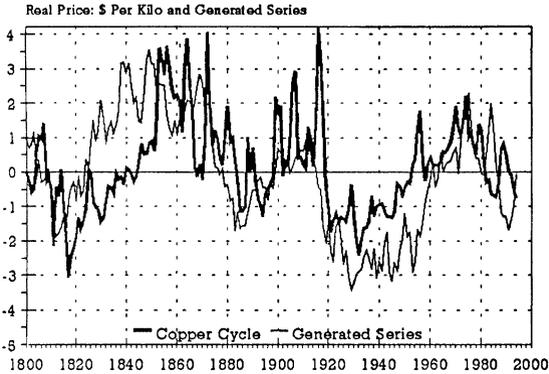


Figure 4(a). – Cycles in real copper prices and a random walk.

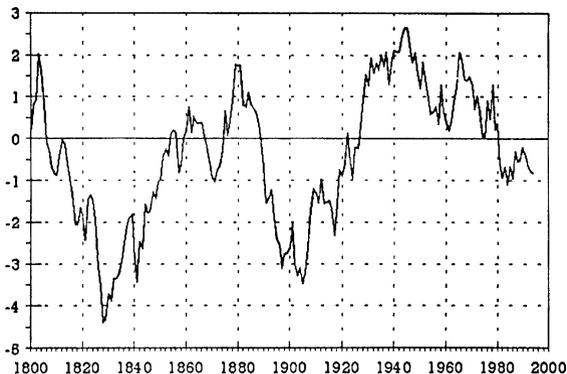


Figure 4(b). – A generated series through a random walk.

Figure 4(a) is the best of ten attempts made to resemble as close as possible the cycle of copper prices (*i.e.*, the deviations of real copper prices shown in Figure 3(d) from their long-term exponential trend). Figures 4(b), 4(c) and 4(d) were selected from among a total of ten tries to illustrate extreme cases of applying equation (3).

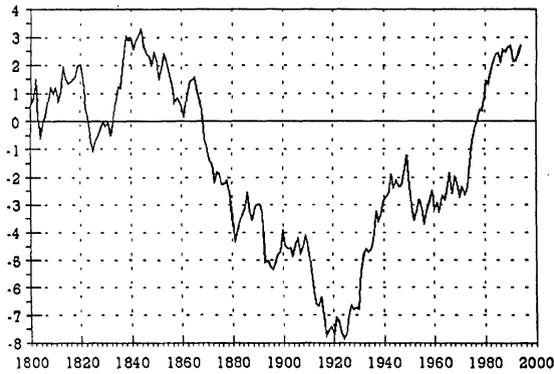


Figure 4(c). - A generated series through a random walk.

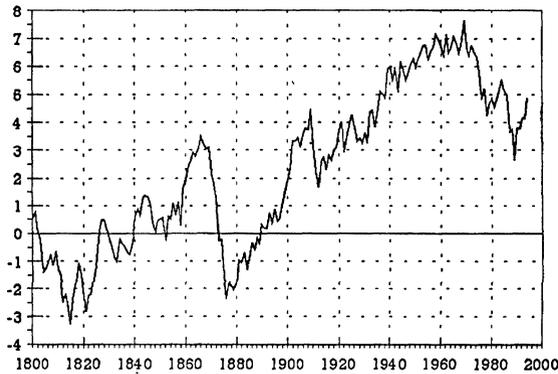


Figure 4(d). - A generated series through a random walk.

The obvious resemblance in Figure 4(a) between the copper price cycle and the series generated through expression (3) is striking, proving that cycles are the outcome of random factors (a fact that has been pointed out by Slutsky, 1937) whose influence is being accumulated over time. The shapes of Figures 4(b), 4(c) and 4(d) suggest that the cumulative effect of random fluctuations can result in illusory patterns that look like cycles and/or trends and make us believe that there are underlying factors behind them.

Y_t as presented by equation (3) is called a random walk and it is characterized by our inability to predict its next turning point. Many economic and business series behave like random walks making the most appropriate forecast, for any future value of Y_t , the latest one available, or

$$Y_{t+i} = Y_t \tag{4}$$

What is important in Figure 4 is to realize the consequences of initial conditions and random disturbances on the final outcome. Figure 4(a) clearly illustrates the effects of chaos on the behavior of an economic variable: copper prices. It suggests that copper prices, once the long-term trend has been excluded, are chaotic, or random walks, and cannot be predicted, unless additional non-quantitative information is available (e. g., inside information about capacity utilization rates, cartels, etc.) which are behaving like random walks. The various cycles shown in Figures 2(b), 3(c) or 4(d) are, in effect, the consequences of cumulative random fluctuations. Figures 2(b), 4(c) and 4(d) suggest that random walks can take forms that may make us believe, wrongly, in the existence of strong patterns which are illusory, however.

Regular/Repetitive generating processes

Some processes are not only repetitive, but also seasonal. A typical case can be seen in Figure 5 which shows the average monthly temperatures in Paris since 1970 or Figure 6 which displays monthly gin sales in the USA for twelve years. Figures 5 and 6 have been generated by the process

$$Y_t = Y_{t-12} + e_t \quad (5)$$

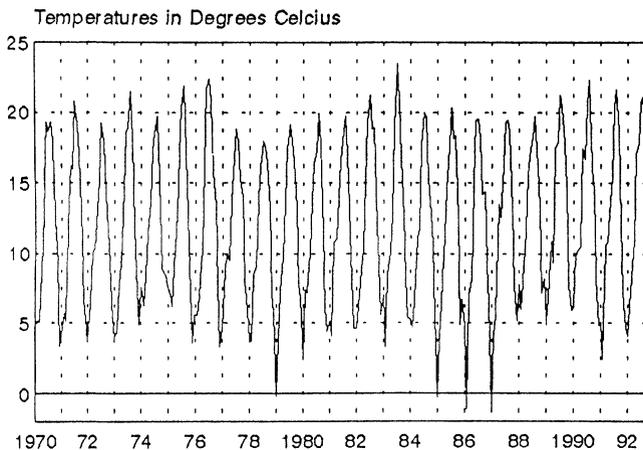


Figure 5. – Average monthly Paris temperatures.

For forecasting purposes predictions involving equation (5) are the most accurate ones as the magnitude of e_t is a small part of Y_t . Moreover, the seasonal part of the generating process stays constant while the cyclical and trend which may be included in (5) exert little influence for the short time

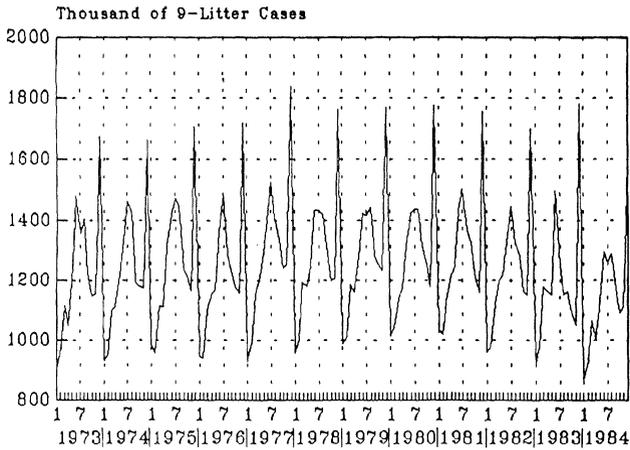


Figure 6. – Gin sales: 1973-1984.

intervals (e. g., months) represented by the series. However, when cycles and/or trends also become influential then future values of Y_t cannot be predicted as accurately as in the case of Figures 5 or 6 for three reasons. First, the influence of e_t is not necessarily additive. Second, a cyclical turning point might affect future values of Y_t , and finally even the trend of the generating process cannot be assured to be constant since in practically all cases the available data will be short. This means that the overall generating process is of the form:

$$Y_t = f(Y_{t-L}, C_t, T_t, e_t) \tag{6}$$

where L is the length of seasonality, 12 for the monthly data shown in Figures 5 and 6, C_t is the cyclical part of the generated process as defined by (3), and T_t is the trend of the generating process.

When (6) applies it is much more difficult to identify the real generating process unless the data series available is much longer than the longest cycle.

Near white noise processes

A large number of series in the economic and business areas, in particular those referring to short time intervals, are white noise or near white noise and can be expressed as:

$$Y_t = \bar{Y} + e_t \tag{7}$$

A typical series generated by equation (7) can be seen in Figure 7, the inventory demand for product RMEUZ3. For forecasting purposes the best prediction for future values of Y_t is \bar{Y} as e_t cannot be predicted.

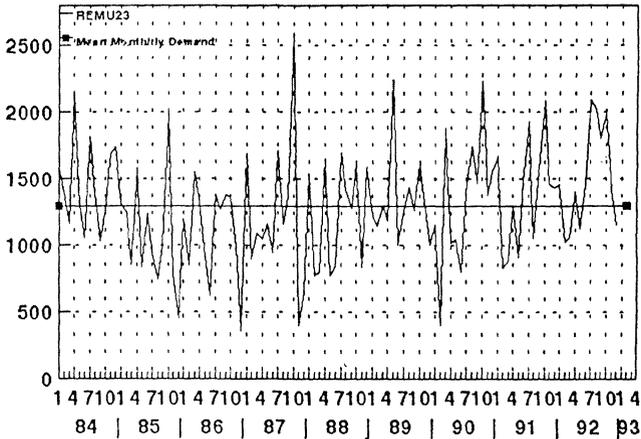


Figure 7. - RMEUZ3: Monthly demand.

FORECASTING AND GENERATING PROCESSES

Although there is an overall generating process, in practical terms we are interested in some aspect of it depending upon the time horizon of interest when forecasting. For instance, for daily or weekly predictions cyclicity is of little importance and trend even less so while seasonality is critical. On the contrary, when long-term forecasts are needed, seasonal fluctuations are of no value while cycles are of little importance. The trend, on the other hand, is critical. We must, therefore, understand the characteristics and behavior of the generating process in relation to the time horizon our forecasts intend to cover. This is more so as people are capable of influencing, or even permanently changing, the generating process through their actions and reactions.

Ability to influence/change the generating process

People are capable of influencing, or even permanently changing generating processes in the economic and business environment. This makes forecasting even more difficult because it is not only possible that we cannot properly identify a certain generating process, or that some other might naturally

change on its own; people can also influence and/or change them. Figure 8 shows the real oil prices since oil was mass produced in 1870. In 1973 oil prices more than doubled and consequently increased more than eightfold going from \$3.4 a barrel in 1972 to \$31.8 in 1981. Such a huge price increase was the direct result of a cartel, OPEC, which was created by oil producing countries and which was successful in controlling oil production. The large increases in oil prices could not have been predicted as it could neither have been known that a cartel would have been created nor that such a cartel would have the cohesion and power to impose production quotas and manage to raise prices to such an extent. However, OPEC's hold could not, and did not last for long. Oil prices fell as OPEC could not maintain its hold of production which kept increasing in spite of all its efforts, even though demand, because of the higher prices, was decreasing. In 1994 the real oil price was about the same as that of 1973 and below the constant long-term trend (see Figure 8).

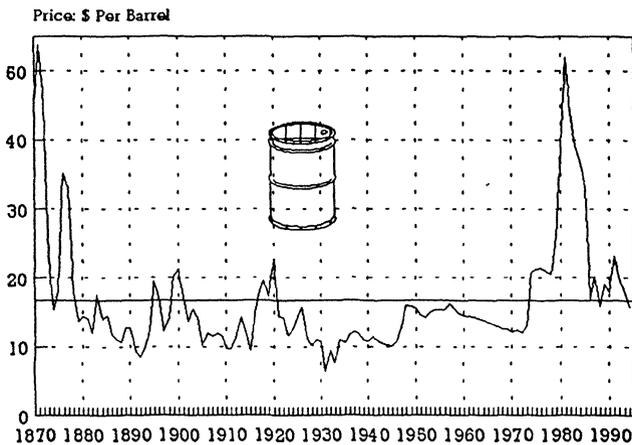


Figure 8. – Oil prices in constant 1994 dollars.

Figure 9 shows gold prices since 1880, the earliest data available. There is a similar huge increase in gold prices starting in 1970 and lasting until 1980, this time not because of a cartel but because gold was used by governments as a reserve for the paper money they were issuing and by individuals as a safe haven against inflation and other instabilities. The use of gold for reserves and as an investment increased its demand and, therefore, price which went from \$36.4 a troy ounce in 1970 to \$612.6 ten years later. The reasons for such a huge increase were, however, purely psychological.

Gold had little other worth than its perceived safety and “agreed” value as a reserve. When governments decided they did not need to keep gold as a reserve, and its perceived safety disappeared, its price fell as fast as it had risen. In 1994 the constant price of gold was the same as that of 1973 and close to its constant long-term trend.

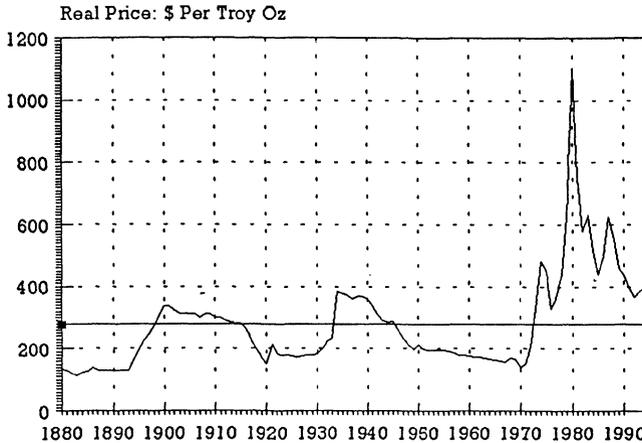


Figure 9. – Gold prices in constant 1994 dollars.

System complexity

We know that the behavior of large, complex systems is chaotic as it is governed by complex interrelationships and multiple causalities. Moreover, we know that such systems are unstable in the short term, as they are strongly influenced by initial conditions, and highly stable in the long run. Large complex systems are governed by a multitude of independent components and are subject to random external influences (Gleick, 1987, p. 303) whose cumulative impact can sometimes contribute to strong short and medium-term fluctuations. What I have shown so far is that business and economic systems are also chaotic and, in addition, can be influenced through human actions, and reactions, making their prediction even more difficult than the corresponding chaotic systems found in the physical or natural environment.

The problem is that business executives, policy makers and others searching for accurate forecasts to improve their decision-making are **not** willing to accept the chaotic behavior of economic/business systems and the extent of uncertainty that surrounds them in the short and medium term. Instead they are looking for substitutes to prophesy which will

eliminate future uncertainty and allow them to achieve deterministic planning. Needless to say, prophecies are neither possible, as economic/business systems are chaotic, nor useful even if they could, somehow, become a realistic alternative. For instance, consider the possibility that a model could predict, with perfect accuracy, which stocks will do well or that a certain product will grow by 20% each year for the next ten years. In such cases the price of the stocks will instantly increase to account for the certainty that it will do well, while high growth forecasts will create strong competition as existing firms and new entrants will increase their capacity to profit for the predicted strong increase in demand. It is practically certain, however, that increased capacity will bring oversupply and strong competition which, in all likelihood, will result in low profits, or even losses. This means that our limits to predictability must be accepted (as it has already been done in physical and natural sciences where chaos is considered as important as order) and the uncertainty surrounding all future predictions acknowledged. After all, we cannot expect to be capable of forecasting social systems any better than meteorological (Lorenz, 1966) or natural ones (Peterson, 1993). As the weather and the motion of planets are governed by chaotic aspects, so are economies, industries and firms.

Long-term trends and various cycles

Figure 10 shows real wheat prices (one of the longest series for which reliable data is available) in England since 1264. There are some definite trends lasting for long periods of time in Figure 10. The first involved a small but persistent decline that went on from 1264 to 1570. The second consisted of an upward trend that lasted from 1570 to the beginning of the nineteenth century. The final trend, starting around 1800, is downward sloping and of an exponential fashion, very similar to that of copper prices shown in Figure 3(d).

This trend, part of Y_t , is

$$Y_t = f(T, u_t) \quad (8)$$

where T is the Trend and u_t is

$$u_t = (\Sigma e_t, \Sigma a_t) \quad (9)$$

where Σe_t is defined by (3) and Σa_t refers to the cumulative effect of human actions to influence future outcome.

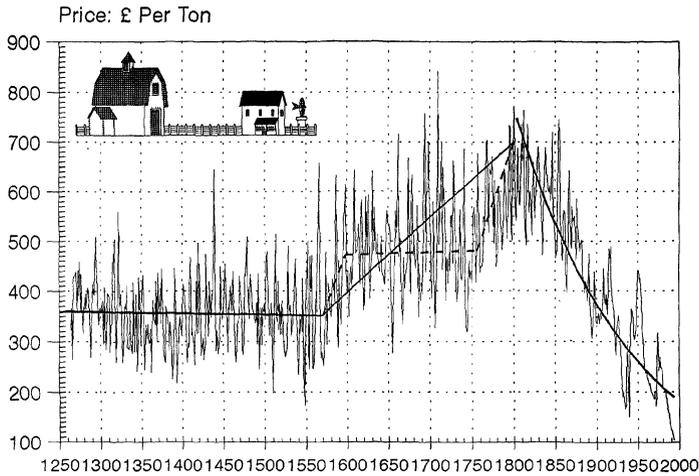


Figure 10. – Wheat prices in constant 1993 £.

It is clear that e_t cannot be predicted as it is made up of white noise while it is debatable whether or not some part of a_t is forecastable as we can observe the cumulative impact of human actions over time. Maybe a good knowledge of the initial conditions of the series we are interested in as well as the cumulative influence of human interventions can contribute to estimating more accurately (8) and predicting forthcoming turning points in equation (9), somewhat better than just using (4), but there is not enough evidence to allow us to support such a statement, although some claim that they can predict turning points more accurately than (4) (that is using today's values as the most accurate forecasts).

If we are sure that the long-term trend is constant then the probability that there will be a regression towards such a trend increases the further away a series gets from the long-term trend. This means that the likelihood of a turning point increases the further away the data gets from their long-term trend. However, it might take many decades before a turning point occurs as it can be seen in Figure 3(c). At the same time trends can and do change as Figure 10 can confirm. We must, therefore, persuade ourselves that the process generating the trend is indeed constant. However, once we can assure ourselves that the trend is constant (as in the copper or wheat prices after 1800 when the effects of the industrial revolution started) then long-term forecasting can be highly accurate.

The biggest problem as far as forecasting is concerned comes in the medium and longer term, when not enough data is available, and where

upward or downward swings whose origin can be cyclical, or psychological can be confused with trends (as in Figures 3(a), 3(b) and 3(c) or with the graphs of Figure 4 which are actually random walks). Such a confusion can provide disastrous forecasts exactly because the likelihood of a turning point increases the longer the “presumed” trend goes away from the real trend of the generating process. For example, consider extrapolating Figure 3(c) and the disastrous forecasts it would have produced as there was indeed a regression towards the long-term trend shown in Figure 3(d), and therefore a turning point which occurred after 1973. The same will be true if we assume, wrongly, that there is an upward trend in Figure 4(c) that started in 1922 as the series is actually a random walk generated by equation (3).

Unfortunately, the great majority of series in the business world are extremely short, barely extending more than four or five years. In such cases it is practically impossible to separate short, medium and long-wave cycles, or other causes, from trends and we have to be content with forecasting the trend **and** cycles together and not believing that changes in trends can be predicted unless we can confirm them with several/many decades of data. Moreover, we cannot expect to predict cyclical turning points as such points cannot be forecast for series generated by equation (3).

Repetitive, seasonal patterns

Like fractals, seasonality repeats itself at regular intervals and can, therefore, be predicted although it is intermixed with trend, cycles and white noise. Moreover, seasonality can be easily identified and measured even when short series, covering a few years, are available. Finally, the process generating seasonality seems constant, or changing very slowly and usually in a predictable fashion, allowing us to extrapolate it, with a reasonable degree of confidence, beyond available data. The major problem is not the incapability of predicting seasonality accurately but rather that cyclical turning points, and other systematic fluctuations (fashions, special events or actions etc) can produce substantial errors, when they occur, which decrease overall accuracy substantially.

Post-sample forecasting accuracies and generating processes

As already mentioned, forecasting competitions have shown that simple methods like single exponential smoothing are extremely accurate because

they assume that the generating process can be described by equation (3) or (7). In such a case their forecasts are optimal as they are of the form

$$F_{t+i} = \alpha X_t + (1 - \alpha) F_t \quad (10)$$

where F_t denotes forecasts.

Equation (10) can optimally predict (7) by using a value for α close to 0, or adequately extrapolate (3) by using a value for α close to 1. Moreover, in case that the generating process is a combination of (3), (5) and/or (7), equation (10) can also be used by determining a value for α between 0 and 1 and estimating and extrapolating the seasonality of the series. More sophisticated methods, on the other hand, attempt to determine a consistent, no changing, trend which they consequently extrapolate assuming constancy. This results in less accurate forecasts than simpler methods such as that described by (10), when, say, a rising cyclical component is taken to be an increasing trend and extrapolated. In other words sophisticated methods are **too** quick to learn and assume that increases or decreases in the data are caused by a constant trend in the generating process which will not change in the future. This assumption is impossible to guarantee from the limited amount of data available which only allows us to study a very small snapshot of the long-term behavior of the generating process.

Unless we possess extremely long series such as those employed in the Santa Fe Competition (Weigend and Gershenfeld, 1994) where the shortest series included 1000 points and the longest 34000, we could never be certain of distinguishing cycles from long-term trends, as in the Paris temperatures series shown in Figure 2, and make strong statements about the trend of the generating process. In the usual case of short series our best choice is to predict the seasonality of our series as accurately as possible and then use expression (10) that accepts that we cannot predict cyclical turning points. Unless we have series, or other information, allowing us to predict the long-term trend we had better only stick to identifying and extrapolating the repetitive, seasonal aspect of the series (the equivalent of fractals) and accept that not much more is possible.

CONCLUSION

Unless we possess an infinite amount of data we cannot estimate the real generating process of series. In particular it is practically impossible to distinguish cyclical swings from trends, make sure that a trend has or is about

to change, or predict when the next cyclical turning point will occur. As the great majority of series coming from the economic and business environment are short (a few to several years of data) it is impossible to use them for forecasting much more than their seasonality unless additional judgmental information is available to us. This means that the short-term behavior of the series can be predicted fairly well by simple methods which can accurately extrapolate seasonality and adequately deal with randomness. This is what forecasting competitions (Fildes and Makridakis, 1995) have shown.

When extremely long series are available (Weigend and Gershenfeld, 1994) it is possible to find sophisticated methods capable of more effectively identifying the real generation process and more accurately forecasting. However, such a task is practically impossible in the economic and business world where having three years of monthly data (*i.e.* 36 observations) is considered a long series (versus up to 34000 observations used in the Santa Fe Competition, Weigend and Gershenfeld, 1994).

Unless much longer series become available and more is learned about the generating processes of businesses and economic events we will have to accept the conclusion of empirical accuracy studies and be content with using simple methods and their combining to forecast series in the business and economic environment which is complex and chaotic. We can assume, as most people do at present, that the world is orderly and we can forecast it with some exceptions when we will be wrong, being surprised when making large errors; or alternatively we can accept reality and realize that the business and economic environment is unpredictable with some exceptions, being realistic. Of course, improvements in forecasting accuracy, through whatever means we can achieve them, can provide us with great value.

REFERENCES

- J. S. ARMSTRONG, Forecasting with Econometric Methods: Folklore versus Fact, *Journal of Business*, 1978, Vol. 51, No. 4.
- G. E. P. BOX and G. JENKINS, *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco, 1970.
- C. CHATFIELD, Calculating interval forecasts, *Journal of Business and Economic Statistics*, in press, 1993.
- R. CLEMEN, Combining forecasts: A review and annotated bibliography with discussion, *International Journal of Forecasting*, 1989, Vol. 5, pp. 559-608.
- K. O. COGGER, The optimality of general-order exponential smoothing, *Operations Research*, 1974, Vol. 22, pp. 858-867.
- F. COLLOPY, and J. S. ARMSTRONG, Rule-based forecasting, *Management Science*, 1992, Vol. 38, pp. 1394-1414.

- J.-C. DUPLESSY, and P. MOREL, *Gros Temps sur la Planète*, Editions Odile Jacob, Paris, 1990.
- R. FERBER, Are correlations any guide to predictive value, *Applied Statistics*, 1956, Vol. 5, pp. 113-122.
- R. FILDES, and S. MAKRIDAKIS, The Impact of Empirical Studies on Time Series Analysis and Forecasting, *International Statistical Review*, forthcoming, 1995.
- R. FILDES, An evaluation of Bayesian forecasting, *Journal of Forecasting*, 1983, Vol. 2, pp. 137-150.
- E. S., Jr. GARDNER and E. MCKENZIE, Forecasting trends in time series, *Management Science*, 1985, Vol. 31, pp. 1237-1246.
- M. D. GEURTS, and J. P. KELLY, Forecasting retail sales using alternative models, *International Journal of Forecasting*, 1986, Vol. 2, pp. 261-272.
- J. GUIOT, J. L. de BEAULIEU, A. PONS and M. REILLE, A 140,000 year climatic reconstruction from two european pollen records, *Nature*, 1989, Vol. 338, pp. 309-313.
- J. GLEICK, *Chaos: Making a New Science*, Viking, New York, 1987.
- W. R. HUSS, Comparative analysis of company forecasts and advanced time series techniques in the electric utility industry, *International Journal of Forecasting*, 1985, Vol. 1, pp. 217-239.
- J. JOUZEL, C. LORIUS, J. R. PETIT, C. GENTHON, N. I. BARGOV, V. M. KOTLYAKOV and V. M. PETROV, Vostok Ice Core: A continuous isotope temperature over the last climatic cycle (160 000 years), *Nature*, 1987, Vol. 329, pp. 403-408.
- A. B. KOEHLER and E. S. MURPHREE, A comparison of results from state space forecasting with forecasts from the Makridakis Competition, *International Journal of Forecasting*, 1988, Vol. 4, pp. 45-55.
- E. LORENZ, Large-Scale Motions of the Atmosphere: Circulation, in *Advances in Earth Science*, in Hurley, P. M. (ed), The MIT Press, Cambridge, Mass, 1966.
- E. J. LUSK and J. S. NEVES, A comparative ARIMA analysis of the 111 series of the Makridakis competition, *Journal of Forecasting*, 1984, Vol. 3, pp. 329-332.
- S. MAKRIDAKIS, C. CHATFIELD, M. HIBON, M. LAWRENCE, T. MILLS, K. ORD and L. F. SIMMONS, The M-2 Competition: A real-life judgmentally based forecasting study, *International Journal of Forecasting*, 1993, Vol. 9, pp. 5-29.
- S. MAKRIDAKIS, Sliding Simulation: A New Approach to Time Series Forecasting, *Management Science*, 1990, Vol. 36, pp. 505-512.
- S. MAKRIDAKIS and R. L. WINKLER, Sampling distributions of post-sample forecasting errors, *Applied Statistics*, 1989, Vol. 38, pp. 331-342.
- S. MAKRIDAKIS, M. HIBON, E. LUSK and M. BELHADJALI, Confidence intervals: an empirical investigation of the series in the M-competition, *International Journal of Forecasting*, 1987, Vol. 3, pp. 489-508.
- S. MAKRIDAKIS, The art and science of forecasting: an assessment and future directions, *International Journal of Forecasting*, 1986, Vol. 2, pp. 15-39.
- S. MAKRIDAKIS, A. ANDERSEN, R. CARBONE, R. FILDES, M. HIBON, R. LEWANDOWSKI, J. NEWTON, P. PARZEN and R. WINKLER, The accuracy of extrapolation (time series) methods: Results of a forecasting competition, *Journal of Forecasting*, 1982, Vol. 1, pp. 111-153.
- S. MAKRIDAKIS and M. HIBON, Accuracy of Forecasting: an empirical investigation with discussion, *Journal of the Royal Statistical Society (A)*, 1979, Vol. 142, pp. 97-145.
- R. MAY, Simple Mathematical Models with Very Complicated Dynamics, *Nature*, 1976, Vol. 261, pp. 459-467.
- R. MAY, Biological Populations with Non-Overlapping Generations: Stable Points, Stable Cycles, and Chaos, *Science*, 1974, Vol. 186, pp. 645-647.

- S. K. McNEES, An Evaluation of Economic Forecasts, *New England Economic Review*, Nov-Dec, 1975.
- D. C. MONTGOMERY and L. A. JOHNSON, *Forecasting and Time Series Analysis*, McGraw-Hill, New York, 1976.
- P. NEWBOLD and C. W. J. GRANGER, Experience with forecasting univariate time-series and the combination of forecasts with discussion, *Journal of the Royal Statistical Society (A)*, 1974, Vol. 137, pp. 131-165.
- P. N. PANT and W. H. STARBUCK, Innocents in the forecast: forecasting and research methods, *Journal of Management*, 1990, Vol. 16, pp. 433-460.
- I. PETERSON, *Newton's Clock: Chaos in the Solar System*, W.H. Freeman and Company, New York, 1993.
- M. B. PRIESTLEY, Comment on Makridakis, Spyros, Hibon, Michèle, Accuracy of forecasting: an empirical investigation with discussion, *Journal of the Royal Statistical Society (A)*, 1979, Vol. 142, pp. 127-128.
- D. J. REID, A comparative study of time series prediction techniques on economic data, *Ph. D. Thesis*, University of Nottingham, Nottingham, 1969.
- S. P. SCHNAARS, A comparison of extrapolation models on yearly sales forecasts, *International Journal of Forecasting*, 1986, Vol. 2, pp. 71-85.
- M. P. SCHUPACK, The predictive accuracy of empirical demand analysis, *Economic Journal*, 1962, Vol. 72, pp. 550-575.
- E. SLUTSKY, The Summation of Random Causes as the Source of Cyclic Processes, *Econometrica*, 1937, Vol. 5, pp. 105-146.
- M. W. WATSON, L. M. PASTUSZEK and E. CODY, Forecasting commercial electricity sales, *Journal of Forecasting*, 1987, Vol. 6, pp. 117-136.
- A. S. WEIGEND, and N. A. GERSHENFELD, *Time Series Prediction: Forecasting the Future and Understanding the Past*, Addison-Wesley Publishing Company, Reading, Mass, 1994.