1987-06

A New Approach to Statistical Forecasting

Makridakis, Spyros

Springer Berlin Heidelberg

"A NEW APPROACH TO STATISTICAL FORECASTING"

by
Spyros MAKRIDAKIS*

N° 87 / 20

* Spyros MAKRIDAKIS, INSEAD, Fontainebleau, France

Director of Publication :

Charles WYPLOSZ, Associate Dean
for Research and Development

Printed at INSEAD,
Fontainebleau, France

A NEW APPROACH TO STATISTICAL FORECASTING

Spyros Makridakis

Research Professor, INSEAD

PRELIMINARY DRAFT: Please do not quote or reproduce without permission

Address
INSEAD
Bld de Constance
77305 Fontainebleau
France

June, 1987

C O M M E N T S

A R E

W E L C O M E

A NEW APPROACH TO STATISTICAL FORECASTING

Abstract

Available approaches to statistical forecasting suffer from several deficiencies (problems) that render their predictions for real-world economic/business series inappropriate. In this paper I illustrate such deficiencies, with real-world data, and propose a new approach that corrects their negative impact. In addition to being theoretically appealing, this new approach outperforms the best method of the M-Competition by a large margin when tested empirically.

A NEW APPROACH TO STATISTICAL FORECASTING

by

Spyros Makridakis

Research Professor, INSEAD

The last ten years have become a learning ground for those working in the field of forecasting. Large-scale empirical studies (Ahlers and Lakonishok, 1983; Makridakis and Hibon, 1979; Makridakis et al. 1982; Zarnowitz, 1984) have provided us with invaluable information to judge the accuracy of statistical methods and to help better understand their advantages and limitations. Few will disagree that the findings of these studies are fundamentally changing the field of forecasting. The fact that simple methods were found to be as accurate as complex or sophisticated ones, and the conclusion that combining methods by a simple arithmetic average did better than the individual methods being combined, brought a great deal of disappointment. The initial disappointment, however, has been replaced by a sense of realism. The alternative of abandoning statistical forecasting does not seem attractive, since the accuracy of judgmental forecasts has been found to be even worse than those of statistical methods (Dawes, 1979; Dawes, 1986; Goldberg, 1970; Hogarth and Makridakis, 1981). It became evident, therefore, that the problems facing statistical forecasting had to be understood and new, imaginative ways of correcting them found (Cogger, forthcoming; Ord, forthcoming; Gardner and McKenzie, 1986; Belsley, 1987; Makridakis, 1987).

In this paper I use several real-world economic/business series to illustrate the deficiencies (problems) surrounding the traditional approach to statistical forecasting and I discuss the reasons for such deficiencies. Second, I propose and empirically test a new approach to deal with these problems. The results I obtain are superior, by a large margin, to the best method of the Makridakis or M-Competition (Makridakis et al., 1982). These results are presented in a way that improvements in forecasting accuracy can be attributed to the various components of the proposed approach and the different selection criteria utilized.

DEFICIENCIES (PROBLEMS) OF STATISTICAL FORECASTING

The most positive outcome of the empirical findings has been the realization and acceptance of the unhappy conclusion that major problems beset the field of statistical forecasting. Problem awareness has brought a gradual but also fundamental change of attitudes to those working in the field. It is now accepted that the traditional approach to statistical forecasting cannot adequately deal with real-world series (Armstrong, 1986; Clement and Winkler, 1986; Mahmoud, 1984; Makridakis et al., 1982; Makridakis, 1986). In this section I summarize what I consider to be the two major deficiencies of the traditional approach, and provide examples to illustrate these deficiencies.

Model Fitting Versus Forecasting

In the traditional approach to statistical forecasting, a model of a method (or methodology) is fitted to all available time series or cross-sectional data. The choice of the method (or methodology) is a matter of personal preferences with some guidelines drawn from previous empirical studies. Once a method (or methodology) has been selected, the specific model that best fits the available data for one-period-ahead forecasts is selected and used to predict for the future (post-sample). "Best fit" commonly means the model that minimizes the Mean Square Error (MSE), the Mean Absolute Percentage Error (MAPE), Mean Absolute Deviation (MAD), Median, Akaike's information, or some analogous criterion (see Exhibit 1). Theoretically, models that minimize m-period-ahead forecasts (by making a period t forecast, $X_{t+m}(m)$, aimed at m-periods later) also exist, but in practice their usage is limited, or nonexistent. This is due to the following three reasons: (a) the theoretical advantages of these models over those making one-period-ahead forecasts, which are then bootstraped to predict for two, three, ..., m periods, are not clear; (b) empirical evidence has shown no real differences in post-sample forecasting accuracy (Makridakis et al., 1982; Andersen and Carbone, 1983) between one and multi-step-ahead forecasts; and (c) available software rarely includes options for multiple-period forecasts. Thus, for practical purposes it can be assumed that only one-period-ahead models are available - or in any event, I am not aware of any exceptions.

In some methods such as ARIMA or Regression, model errors or the random disturbances need to be independent, constant and normally distributed. In other methods (exponential smoothing, Bayesian forecasting) there is no restriction about the disturbances although it is desirable that they be random, constant and normally distributed. None of the methods are concerned or can know the properties of the post-sample forecasting errors, which are assumed to posess properties analogous to those of the model's disturbances, although this is not ordinarily true (Makridakis and Winkler, 1985).

Two assumptions are implicit in the traditional 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 the great majority of real-world economics/business series. Exhibit 2, for instance, shows the ranking of five methods for 1,2,3,..., 18 forecasting horizons at different chronological time periods. That is, the best (denoted by 1), second best (denoted by 2), ..., worst (denoted by 5) method at each time period (starting from 67 and going to 127) was found for 1,2,3,..., and 18 period-ahead forecasts. Such rankings are not consistent at different time periods or forecasting horizons. The series in Exhibit 2 is typical of those of the M-Competition.

The same conclusion can be drawn by computing the rank correlation between how well the methods in the M-Competition fitted past data versus how well they forecast beyond these data. Such rank correlations were small to start with (about 0.20) and dropped to zero after forecasting horizons of longer than four periods (Makridakis, 1986).

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 only recently been considered (Priestley, 1979). Even during the 1970's the latter possiblity it was not mentioned in the most popular forecasting or econometric textbooks (Box and Jenkins, 1970; Johnston, 1972). Furthermore, no serious effort was made to validate the ability of the selected model to accurately forecast for out-of-sample periods. This was partly because all data were 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 as much as, series used in the social sciences are short, measurement errors abound, and controlled experimentation is not possible, the basic premise that the "best" model fitted to past data exists and can be identified, and that such a model is also the best model to forecast beyond these data is invalid.

Constant Versus Changing Patterns/Relationships

Exhibit 3 shows monthly international airline passengers (in thousands) between 1949 and 1981. The data is divided into three parts. Part (a) consists of 144 observations (1949-1961), which are the infamous airline data widely used in the forecasting literature since the early 1960's (Box and Jenkins, 1970; Brown, 1963). Part (b) includes data from 1961-1967, while the data in Part (c) are after 1967. There is an obvious change in the pattern of international airline passengers after 1967 (Part (c)). Both the exponential trend and the seasonal fluctuations are different from those existing before 1967. The traditional approach to statistical forecasting assumes constancy of patterns and/or relationships. Such an assumption permits the use of the best model fitted to available data to forecast beyond these data. Unfortunately, however, constancy is not a realistic assumption (Makridakis, 1981 and 1986) as far as business or economic data are concerned, which raises a major issue about the validity of the traditional approach to statistical forecasting.

Since constancy of pattern/relationships is a prerequisite of the traditional approach, "nice" series such as the airline data (part (a) of Exhibit 3) had to be found to test new forecasting methods and illustrate their alleged "superior" performance. Furthermore, it was considered normal to test a new method on a single series (such as the airline data) and then generalize that the same accuracy level would hold for any other series. Having followed this practice myself (Makridakis and Wheelwright, 1977), I can now say that from a methodological point of view it is ludicrous to consider it possible to generalize from the past to the future (or from available data to post-sample), and from a single series to all series. As part (c) of Exhibit 3 shows this is not even possible for what seemed to be a perfect series back in the 1960's.

In addition to patterns, relationships can and do change. Exhibit 4 is the scatter diagram between paper orders in France and pulp prices. At least four relationships can be identified (A,B,C and D), as well as two cases (E and F) where pulp price increases did not affect (reduce) orders. Assuming that the relationship between price and orders is constant is not realistic and results in inaccurate forecasts. Econometricians might argue that the factors causing the relationship between paper orders and prices to shift could be found (if no factors can be found they include dummy variables). Although in some cases this might be possible, it cannot help to forecast more accurately (although the R^2 of the model fit will be better) since the majority of the factors causing the relationship to change are exogenous and, therefore, unpredictable themselves.

Although the majority of forecasting methods would provide equally-good forecasts for data series when there are no changes in established patterns and/or relationships (see Exhibit 5), the forecasts and their accuracy will vary substantially when changes in patterns and/or relationships occur. It is necessary, therefore, to understand how various methods forecast when such changes do take place, since this is the key to understanding the deficiencies (problems) of available methods and to becoming capable of forecasting in the real world when constancy of pattern/relationships cannot be assured. Although changes of relationships also need to be considered, in the remainder of this paper I concentrate on the effects of pattern changes on forecasting.

Exhibits 6, 7, and 8 (the data of the three Exhibits have been deseasonalized to better illustrate pattern changes and their consequences) show three kinds of pattern change during forecasting. In Exhibit 6 the exponential-growth trend changed into an abrupt decline. There was nothing in the past data to indicate that such a change was forthcoming. It was impossible, therefore, to have anticipated a pattern change without exogenous judgmental knowledge. All methods, except for single exponential smoothing, forecast a continuation of the established trend (single exponential smoothing always forecasts horizontally at the most recent smoothed data level). Contrary to the data of Exhibit 5 where the trend continued and single exponential smoothing did not forecast accurately because it assumed no trend, in the case of Exhibit 5 exponential smoothing performs the best, since all methods (except linear trend regression) forecast by extrapolating the established exponential trend.

The data of Exhibit 7 start increasing at period 34 and do so until period 39. The figures then decrease for two consecutive periods. Two methods (Box-Jenkins and quadratic smoothing) ignore the latest two-period decline and forecast a continuation of the recent increase from periods 34 to 39. Bayesian Forecasting assumes that the decline in period 39 and 40 is not random, and forecasts by extrapolating the downward trend implicitly assuming the latest decline to be permanent. By so doing, the Bayesian procedure produces forecasts that beat all other methods. Linear trend regression ignores all fluctuations around the trend line, assuming them to be random, and extrapolates the trend to arrive at linearly-growing forecasts. The forecasts of the other methods are between those of regression and Box-Jenkins. Interestingly, single exponential smoothing does pretty well, although it ignores both the initial increase (periods 34 to 39) in the actual data and the subsequent two-period decline (see Exhibit 7). The series in Exhibit 7, contrary to that of series 6, has in its past provided indications that it might decline after several periods of continuous increase. Such a decline has happened twice in the past. One could therefore have anticipated that a similar decline might occur in the future and have forecast in this light.

The data of Exhibit 8 reach a trough at period 96, then they increase (with small interuptions) until period 120 at which point they start declining until period 125. Finally, there is a single increase at period 126. Bayesian Forecasting, although doing best with the data of Exhibit 7, does the worst with those of Exhibit 8. It assumes a growing trend, thus providing increasing forecasts. Quadratic exponential smoothing which did the worst with the data of Exhibit 7, now does the best by ignoring the increase in the last period and forecasting a continuing decline from periods 120 onwards. The forecasts of the other methods are in between those of quadratic smoothing and Bayesian Forecasting. The series of Exhibit 8 is similar to that of Exhibit 7 in that several declines after persistent increases, similar to the latest one, have occurred in the past. It is not unreasonable, therefore, to anticipate (although the exact timing might not be predictable) that similar declines might occur in the future during forecasting.

Three observations are worth making at this point. First, the forecasts of the various methods are all over the graph when a pattern change occurs (see Exhibit 6, 7, and 8) during the forecast period (this is one reason why combining various methods by simple arithmetic averaging does well). Second, the accuracy of the methods depends upon whether the latest non-random change in pattern is temporary or permanent (Makridakis, 1986).

Some methods, such as Bayesian Forecasting, are reactive in extrapolating recent non-random changes in the data pattern by assuming them to be permanent. Other methods are slower in identifying and extrapolating the continuation of non-random changes in the data. Linear trend regression, for instance, ignores all changes around the long-term trend, while single exponential smoothing assumes a no-change (trend) situation. Third, single exponential smoothing seems to do well, not because it can predict pattern changes, but rather because its forecasts are robust, staying in the middle of the data and usually being in the middle of the forecasts of the various methods when patterns change. This seems to be a good strategy, at least for the short term, since empirically the accuracy of single exponential smoothing for one-period-ahead forecasts was found to be the best of all methods in the M-Competition. Moreover, it seems to work well also for longer forecasting horizons, since its forecasts seem to be accurate.

THE PROPOSED APPROACH

For any forecasting approach to be realistic and practically relevant it must avoid the two major problems facing the traditional approach to statistical forecasting - that is selection based upon how well a model fits available data for one-step-ahead forecasts, and assuming constancy of patterns/relationships. In addition, it needs to incorporate what we have learned from empirical studies (see Exhibit 9), and it must permit one to test forecasting performance on out-of-sample data.

Initially, the desired characteristics of the new approach might seem contradictory. Any time series model, for instance, must be based on past data. At the same time, I contend that the future might not be the same as the past. Furthermore, all data should be used to develop the forecasting model (otherwise some information might be lost), while at the same time I advocate that out-of-sample testing needs to be done. These seemingly contradictory requirements can be simultaneously achieved if we are willing to reconceptualize our approach to statistical modelling and forecasting.

Exhibit 10 shows an approach to model selection that is a viable alternative to that of Exhibit 1. Instead of using all n data points to develop a forecasting model, only s<n data points are initially used and m forecasts are made. Since actual data exist beyond s, the actual forecasting accuracy of the model can be tested for each of the m forecasts. Accuracy measures (such as MAPE, MSE or Median) for 1,2,3,..., m-period-ahead forecasts can, therefore, be found. Subsequently, one more data point can be used, m forecasts made, and their actual forecasting accuracy recorded. The process can be done, each time using one more data point, until all observations except one have been used. This type of testing (simulation) I call out-of-sample and is shown schematically in Exhibit 10.

Once the Jackknife simulation has been completed k one-step-ahead accuracy measures, k-1 two-step-ahead accuracy measures (k=n-s, where n is the number of data points, and s is the starting period before the simulation starts) , ..., k-m+1 for m-step-ahead accuracy measures are available. The average of these measures for each of the m forecasting horizons can be computed and the average of these m averages can also be found, if so desired (see Exhibit 10). Subsequent model selection can be based on actual out-of-sample forecasting performance without any loss of information, since in the final analysis all data have been used. Such a type of model selection is fundamentally different from the traditional approach in two respects. First, model selection is based on forecasts of out-of-sample data. Second, forecasting performance is measured, in addition to one period ahead, for two, three, ..., and m-step-ahead forecasts. Once a model has been selected for each of m forecasting horizons based on its out-of-sample performance, it can then be used to predict for the future, that is for making post-sample forecasts, for the specific period(s) it is the "best".

Among-Methods Model Selection

Several authors (e.g., see Jenkins, 1982) have correctly pointed out that combining forecasts makes no sense from a theoretical point of view. Yet, the empirical evidence showing a "consensus" forecast to outperform the individual methods being combined is indisputable (Clemen and Winkler, 1986; Gupta and Wilton, 1987; Mahmoud, 1984; Makridakis and Winkler, 1984). Some of the reasons contributing to the more accurate performance of combining over individual methods can be deduced from Exhibits 5, 6, 7, and 8. In Exhibit 5, for instance, since all methods do well, combining them will do equally well. In Exhibits 6, 7, and 8, the forecasts of the different methods vary widely, thus making their average robust and closer to the center of the unpredictably changing pattern. This average, therefore, provides not only more accurate forecasts, but also forecasts with smaller variance (Makridakis and Winkler, 1983).

There is another fundamental reason why combining works well with real data series, one that relates to our concept of what constitutes the "best" model to represent reality. In the frictionless physical/natural sciences a best model might exist. This is hardly the case, however, in the friction-filled business/economic fields where the "best" model will be different from series to series (see Exhibits 6, 7, and 8), and where the best can vary at each period and with each forecasting horizon (see Exhibit 2). Under such circumstances, is there a better alternative for improving forecasting accuracy to combining different methods and/or models?

Although, additional research might be required to decide what methods to include, I propose five non-seasonal methods to be used in parallel, with the ultimate selection to be made from the "best" among them. The basis for proposing these five methods are the findings of empirical studies (see Exhibit 9). In addition, the characteristics of these five methods are the following: (1) they are simple; (2) their forecasts are intuitive; (3) they can produce forecasts and confidence intervals in an automatic, push-button manner; and (4) they are complementary, specifically - (a) single exponential smoothing assumes that changes cannot be predicted, (b) Holt's exponential smoothing extrapolates a linear trend (weighting more heavily recent data), (c) dampen-trend exponential smoothing (Reference) is similar to Holt's except, as its name implies, it dampens the trend for longer forecasting horizons (d) Brown's quadratic exponential smoothing extrapolates a quadratic trend (weighting more heavily recent data), and (e) Long-term trend (a long non-stationary Auto-Regressive (AR) model similar to the long-memory ARAR models proposed by Parzen, 1982).

During the Jackknife simulation the aforementioned five methods are run in parallel, using an optimal model from each of the five methods that minimizes some error-selection criterion for each. Although, <u>within</u>-method model selection is discussed below, for the moment assume that the "best" model for each of the five methods is selected by minimizing the MSE at each period of the simulation. Thus, the model that minimizes the one-step-ahead model-fitting errors for the first s data points is selected and m forecasts are made. Then, the first s+1 data points are used, the "best" model is found, and m forecasts are made; and so on until <u>all</u> data points except one have been used. The process allows us to compute m forecasts (based on the "best" model as defined by the traditional approach to statistical forecasting) <u>at each period</u> of the simulation, compute MSE, MAPE etc., measures at each step of the simulation, and find the average of these MSE, MAPE for each of m forecasting horizons (see Exhibit 10). There will be k one-period-ahead forecasts, k-1 two-period-ahead forecasts, ..., k-m+1 m-period ahead forecasts (see Exhibit 10) whose average accuracy can be found.

Unlike the traditional approach, a different method can be selected for each series and each of the m forecasting horizons, depending upon the out-of-sample performance of each method during the Jackknife simulation. Furthermore, confidence intervals based on actual forecasting errors can be constructed (see Williams and Goodman, 1971; Makridakis and Hibon, 1987). In addition, we obtain information (i.e., standard errors) about the sampling variation of the accuracy measure with which we are concerned (e.g., the MSE or MAPE), since the values of such measures for each of the k simulations is known. Knowing the empirical sampling distribution can provide us with invaluable information not available through the traditional approach to statistical forecasting (or statistical modelling in general) which can permit us to select the "best" method among the five using criteria other than the smallest MSE, MAPE or Median.

Exhibit 11 shows the results of the forecasting simulation for the series in Exhibit 8 using the five methods mentioned above. Although quadratic exponential smoothing gives the best MSE for one-period-ahead forecasts for model fitting, it is the worst method for more than two-periods-ahead out-of-sample forecasts. The reason is that the series of Exhibit 8 is cyclical which causes the long-term forecasts of quadratic smoothing to miss the actual values by a large amount when a cyclical turning point occurs. The errors in such cases are huge, thus making the performance of quadratic smoothing for out-of-sample forecasts the worst. For one-period-ahead forecasts (both for the model fitting and the out-of-sample predictions), however, the errors of quadratic smoothing are similar to those of other methods, since even during turning points the one-period-ahead forecasting errors of quadratic smoothing are not large.

Exhibit 12 shows the forecasting performance of the proposed approach, together with the results of the most accurate/important methods of the M-Competition. The model selection for each of the five methods was done using the traditional approach of choosing the model that minimizes the one-step-ahead fitting errors. This type of model selection was done to facilitate comparisons. The first part shows the individual accuracy of the five methods used in parallel during the simulation, as well as the accuracy of their combined forecasts. The second part lists the accuracy of the remaining most accurate/important methods of the M-Competition. The remainder of Exhibit 12 shows the accuracy of the five methods used when the selection among the five methods was done with out-of-sample information. Several selection criteria for choosing one of the five methods for each series <u>and</u> for each of the m forecasting horizon were tried. The effect of combining the best two, three or four methods is also shown.

For each of the five methods the model was found that minimized the MSE when fitted to past data. Subsequently, the method among the five with the smallest MSE and MAPE of model fitting was selected to forecast for m periods ahead (this is the closest one can get to the traditional approach in terms of model selection). The MSE and MAPE of such selection although the best for model fitting they are worse than the combining of the five methods for all horizons in the post-sample forecasts. This selection procedure assumes that the model providing the best fit to past data will produce the best forecasts for the future. The evidence in Exhibit 12 shows that it is preferable to combine the forecasts of the five methods rather than attempt to choose the best among them based on past, model fitting performance.

There is an improvement if, instead of using the MSE or MAPE of model fit, the average MSE or MAPE of all (m) forecasting horizons from the simulation is used as the selection criterion. This selection assumes that there is a single best method independent of forecasting horizons. The most dramatic increase, however, occurs when the method with the best MSE, MAPE or Median for each of the m-forecasting horizons is selected to forecast for that specific horizon. The MAPE's using this selection scheme improve substantially, beating the best method in the competition (Parzen's ARARMA models) by 1.6% (this is more than the improvement of Parzen's over single exponential smoothing).

Combining the best two methods at each forecasting horizon further improves accuracy, while combining the best three methods gives the best results. These results seem equally accurate for short as well as long forecasting horizons. Additional selection criteria have been attempted and appear to be promising. For instance, choosing and combining the method(s) with the best MSE, MAPE and RANK gives an overall MAPE of 13.7% (if the same method is, say, the best in MSE and MAPE a weight of 2/3 is given to that method, while if a single method is best in all three criteria its forecast is used exclusively; otherwise, the forecasts of the method with the best MSE, MAPE and Rank are each weighted by 1/3).

An overall MAPE of 13.7% is also found by choosing all methods whose MAPE is within the range of the MAPE of the best method plus the error bound that is found through taking into account the sampling variation (as measured by the standard error) of the MAPE's of the methods involved. Another selection criterion used was to choose a method that prior experience has indicated to be the best (Best PRIOR) and to retain it unless evidence shows that another method produces forecasts which are statistically more accurate; that is, there is statistical evidence to reject the null hypotheses postulating that method A is not the best. Several other selection criteria were utilized and their results can be seen in Exhibit 12. It seems that differences in forecasting accuracy are small and influenced little by the specific selection criterion being used. The important factor is the selection of a "best" method(s) among the five based on out-of-sample accuracy performance.

Whithin-Method Model Selection

Given a particular method, some optimization criterion can be used to select an appropriate model. In the case of the five methods used, this means finding seasonal indicies and optimal parameter values. This selection can be done two different ways. The first requires using the first s data points and finding the model that best fits these data points. This "best" model can be subsequently used to make m forecasts and compute various errors measured in the Jackknife simulation. Then, the first s_1 (s_1=s+1) data points can be used to find the model that best fits this augmented set and make m new forecasts. This optimization process can continue, each time computing the optimal model until all but one of the data points have been used in the model fitting process. Alternatively, the "best" model can be found by using <u>all</u> n data points and then this model, once found, utilized to make m forecasts with s data points, s_1 data points, ..., and so on until all the data points but one have been utilized in the simulation.

Originally, the first approach was used. Then I decided to compute the seasonal indicies using all data points (finding seasonal indicies requires at least 3-4 years of data which put a serious constraint on the starting period of the simulation) and still optimize the model used at each step of the simulation. The results were compatible. Later on, the parameters of the "best" model for all n data points were found and used at each step of the simulation without having to reoptimize the model parameters at each step. To my surprise overall post-sample forecasting accuracy did not change. I, therefore, decided to use the second approach which involved considerably fewer computations. The results of Exhibit 12 are based on such approach to within-methods model selection.

Using the best model found through the second approach the average MSE (MAPE of whatever other criterion) can be computed for each of m forecasting horizons at each step of the simulation. Although the model selection procedure followed is the same as the traditional approach, it allows us to know the actual forecasting performance of the optimally-selected model for each of m forecasting horizons. Such performance can subsequently help us choose the best of the five methods, as described above, by comparing their MSE (MAPE or other criterion).

An alternative approach to within method model selection is to choose the optimal model, based not on the past data, but on out-of-sample forecasting performance. This means using the first s data points , making m forecasts and recording the errors (say MSE's or MAPE's) for each of these m forecasts. Then, the first s_1 (where s_1=s+1) data points are used to make m forecasts and obtain their MSE's or MAPE's. This procedure would continue with s_2 (s_2= s_1+1), s_3, ..., n-1 data points. If this simulation starting with s and ending at n-1 data points, is repeated using alternative models of the same method, the model that optimizes one-period-ahead forecasts can be selected, as can the model that optimizes two-periods-ahead, three-periods-ahead ..., and m-period-ahead forecasts. The approach just described is different from minimizing a multiple-period-ahead forecasting model (i.e., minimizing the errors $e_t = X_t - \hat{X}_{t-m}(m)$, where t=m+1, m+2, ..., n). In the proposed within-method model selection, a one-step-ahead model is fitted to s, s_1, s_2, ..., n-1 data points and m forecasts made at each time period of the simulation. Although the specific model is fitted to past data, the selection is done on out-of-sample forecasting performance. This minimizes the chance of overfitting and provides a more realistic procedure of measuring accuracy, since the fitted model is tested on out-of-sample data, which is not the case with the traditional multiple-period-ahead forecasting models.

Exhibit 13 shows the results of the traditional approach to model selection based on minimizing one-period-ahead MSE for the model fitting, and that of model selection based on out-of-sample MSE performance. Although the results for a one-period-ahead forecasting horizon are similar between the two approaches, the improvement of the proposed out-of-sample within-method model selection starts at forecasting horizon two and becomes larger as the length of the forecasting horizons increases. The improvement is considerable. The accuracy of single exponential smoothing with the new approach is almost identical to that of Parzen's ARARMA models, the best method of the M-Competition. Furthermore, if yearly data (for which single exponential smoothing is inappropriate since it assumes no trend) are excluded the performance of single exponential smoothing is better than that of ARARMA models. The accuracy of ARARMA models, or any other method, will probably also improve if the model selection is based on out-of-sample data for <u>each</u> of the m forecasting horizons. At present, however, it seems equally advantageous to use a single method and select the best model using the proposed approach rather than the best method currently available.

DISCUSSION AND DIRECTIONS FOR FUTURE RESEARCH

The approach to statistical forecasting I am proposing makes theoretical sense (Saaty, 1980). Equally important, when tested empirically with the M-Competition data, it provides superior results in terms of improved forecasting accuracy. I strongly believe, therefore, that what is now required is more academic research and usage of such an approach in applied settings.

In the last five years countless hours of CPU time have been used at INSEAD's computers to come up with new ways of forecasting more accurately. Innumerable decision rules, combining procedures, error measures (MSE, MAPE, Median, MAD, Geometric mean, etc.), new methods, classification schemes (e.g., use method A for macro data, method B for micro data, method C for industry, or method A for monthly, B for quarterly, and C for yearly data), and method and model decision rules were attempted. As in similar work reported in the literature (Schnaars, 1986) any gains in post-sample accuracy were found to be marginal. Once the best method, however, among several run in parallel, was selected based on out-of-sample accuracy measures, important gains in post-sample accuracy were observed. These initial gains were further improved to their present level, are shown in Exhibit 12. Furthermore, when the out-of-sample within-method model selection (see Exhibit 13) was used in conjunction with among-methods selection, the results improved to an even greater extent (see Exhibit 14). In my opinion, the gains in terms of improved accuracy point towards a breakthrough that has important theoretical and practical consequences. My hope is that replications and more research will provide additional insights to contribute to a better theoretical foundation for statistical forecasting (Duong, 1984) and to even greater improvements in forecasting accuracy.

Jackknife simulation provides additional possibilities beyond improved forecasting accuracy. First, realistic confidence intervals can be built for each of the m forecasting horizons. Such intervals need not be symmetric since information about underestimates as well as overestimates around the most likely out-of-sample forecasts is collected. In addition, through an analysis of extreme errors it is possible to warn forecasting users about <u>unusual</u> errors and help them think of ways to prepare to face such errors.

I believe that additional improvements in forecasting accuracy are possible by the appropriate choice of the methods to be run in parallel. Moreover, I think that the forecasts of advanced methods might prove to be superior to those of simple ones, if the best model of such methods is selected based on out-of-sample performance. In addition, optimal decision rules for combining methods based on out-of-sample forecasting errors might further improve forecasting accuracy. These and similar issues need to be investigated through additional theoretical and empirical research. I believe that some new vistas of forecasting research have opened, with critical implications for improving forecasting accuracy and effectiveness. In addition there is a need to consider the implications of the proposed approach on explanatory models to better understand and deal with the effect of changes in relationships during forecasting.

A major concern (and direction for future research) must be towards more effective ways of anticipating changes in patterns and/or relationships. Judgmental information might be critical in this respect and ways of eliciting these judgments, quantifying them, and incorporating them into the final forecasts, in a Bayesian fashion, need to be devised. How can we deal with cyclical changes for instance? Is there a way of predicting cyclical (or other) turning points that improves upon a random "guess"? A purely statistical analysis of cycles provides little information to help anticipate turning points. The more a series deviates from its long-term trend (assuming that such a trend has not changed in a permanent way), however, the greater the chance of a turning point (regression) towards the long-term trend. This is clear in Exhibits 7 and 8. Nevertheless, there are many false turns (W-type troughs or peaks) that make turning-point forecasting inaccurate.

Exhibit 15 shows two series that move similarly in Part A. Nevertheless, any thought of series A being a leading indicator of series B evaporates if we look at Part B of Exhibit 15. Series B is the trend-cycle (the deseasonalized values when randomness has been removed) of writing and printing paper in France (scaled to be printed alongside Series A). Series A is the monthly average of the S&P 500. While important to observe the similarity in the two series, what makes matters interesting is that it is well known (Reference) that the S&P 500 Series cannot be predicted. Is it then possible that Series B also cannot be predicted? As a matter fact, the autocorrelation of the first difference of the deseasonalized values of Series B is small (equal to -.39 for one time lag to become zero for more than one time lag - suggesting inability to predict cyclical turns). Given the similarity of Series B to A and the small or zero autocorrelations there are serious limits to purely statistical predictability that we must explore and accept, since a great many series in the business/economic fields are cyclical.

I believe it is possible to predict cyclical turns, but much more work is needed and a concerted effort between preparers and users of forecasts is required. In my experience I have found that business executives have a deep knowledge and experience of the market and their customers. The difficulty lies in translating such knowledge and experience into forecasts, which is where a great deal of work is required; I expect to see such work taking place in the near future. We must find effective ways of combining judgment together with statistical forecasts, since these two approaches to predicting the future are complementary (see Makridakis, 1987).

Conclusions

In this paper I have proposed and tested a new approach to statistical forecasting. Such an approach aims at eliminating the deficiencies (problems) of the traditional approach to statistical forecasting. It is based on the principle that model selection must be done on actual, out-of-sample forecasting performance. Such selection is made on two levels. First, the best model (within-method selection) of a single method is chosen. Second, the best method among several, run in parallel, is selected. Both the within-method and among-methods selection is done on out-of-sample comparisons. We do not assume that there exist a unique method that can forecasts best for all series and forecasting horizons. This means that a different model/method can be selected for various series, and for each forecasting horizon, based on the actual out-of-sample performance of the method/model for this specific series/forecasting horizon.

The empirical testing of the proposed approach shows large improvements in forecasting accuracy. Such improvements extend to both short, medium and long forecasting horizons, different types of data (yearly, quarterly, monthly), and other classifications. Finally, the improvements come both within method, when the best model is chosen based on out-of-sample information (see Exhibit 13), and among methods, when the best method is chosen based on out-of-sample performance (see Exhibit 12). Combining the within <u>and</u> among-methods selection further improves post-sample forecasting accuracy (Exhibit 14), suggesting a breakthrough over available methods.

References

Ahlers, D. and J. Lakonishok, 1983, "A study of economists' consensus forecasts", Management Science, 29, 1113-1125.

Andersen, A., and R. Carbone, 1983, "One-step versus multiple-steps forecasts", Université Laval Working Paper.

Armstrong, J.S., 1986, "Research on forecasting: A quarter-centry review, 1960-1984", Interfaces, Vol. 16, N° 1, pp. 89-109.

Belsley, D.A., 1987, "Modelling and forecasting reliability", paper presented at the 7th International Symposium on Forecasting, Boston.

Box, G.E.P. and G.M. Jenkins, 1970, Time series analysis: Forecasting and control (Holden-Day, San Francisco, CA).

Brown, R.G., 1963, Smoothing, forecasting and prediction of disctrete time series (Prentice-Hall, Englewood Cliffs, NJ).

Clemen, R.T. and R.L. Winkler, 1986, "Combining Economic Forecasts", Journal of Business and Economic Statistics, Vol. 4, N° 1, pp. 39-46.

Cogger, K., "Future directions in forecasting research" International Journal of Forecasting, forthcoming.

Dawes, R.M., 1979, "The robust beauty of improper linear models in decision making", American Psychologist, 34, N° 7, pp. 571-582.

Dawes, R.M., 1986, "Forecasting one's own preferences", International Journal of Forecasting, 2, pp. 5-14.

Duong, Q.P., 1984, "On the choice of the order of autoregressive models: A ranking and selection approach", Journal of Time Series Analysis, Vol. 5, N° 3, pp. 145-157.

Gardner, E.S. and E. McKenzie, 1985, "Forecasting trends in time series", Management Science, Vol. 31, N° 10, pp. 1237-1246.

Goldberg, L.R., 1970, "Man versus model of man: A rationale, plus some evidence for a method of improving on clinical inferences", Psychological Bulletin, 73, N° 6, pp. 422-432.

Gupta, S. and P.C. Wilton, 1987, "Combination of forecasts: An extension", Management Science, Vol. 33, N° 3, pp. 356-373.

Hogarth, R.M., and S. Makridakis, 1981, "Forecasting and planning: An evaluation", Management Science, 27, N° 2, pp. 115-138.

Jenkins, G.M., 1982, "Some practical aspects of forecasting in organizations", Journal of Forecasting, Vol. 1, N° 1, pp. 3-23.

Johnston, J., Economic Methods, (2nd Ed.), 1972, McGraw-Hill, New York.

Mahmoud, E., 1984, "Accuracy in forecasting: A survey", Journal of Forecasting, Vol. 3, pp. 139-159.

Makridakis, S., 1981, "Forecasting accuracy and the assumption of accuracy", OMEGA, 9, N° 3, pp. 307-311.

Makridakis, S., 1986, "The art and science of forecasting: An assessment and future directions", International Journal of Forecasting, Vol. 2, pp. 15-39.

Makridakis, S., 1987, "Metaforecasting: Ways of improving forecasting accuracy and usefulness" Keynote address presented at the 7th International Symposium on Forecasting, Boston (INSEAD Working Paper N° 87/15).

Makridakis, S. and S. Wheelwright, 1977, "Adaptive filering: an integrated moving average filter for autoregressive time series forecasting", Operational Research Quarterly, Vol. 28, pp. 877-888.

Makridakis, S. and M. Hibon, 1979, "Accuracy of forecasting: An empirical investigation", Journal of the Royal Statistical Society, 142, part 2. pp. 97-125.

Makridakis, S. et al., 1982, "The accuracy of extrapolation (time series) methods: Results of a forecasting competition, Journal of Forecasting, 1, pp. 111-153.

Makridakis, S. and R. Winkler, 1983, "Averages of forecasts: Some empirical results", Management Science, 29, N°9, pp. 987-996.

Makridakis, S. and R. Winkler, 1985, "Sampling distribution of post-sample forecasting errors", INSEAD Research Paper, N° 85/09.

Makridakis, S. and M. Hibon, 1987, "Confidence intervals: An empirical investigation for the series in the M-Competition", International Journal of Forecasting, Vol. 3, forthcoming.

Ord, K.J., "Future developments in forecasting: The time series connection", International Journal of Forecasting, forthcoming.

Parzen, E., 1982, "ARARMA models for time series analyses and forecasting", Journal of Forecasting, Vol. 1, pp. 67-82.

Priestley, M.B., 1979, "Discussion of the paper by Professor Makridakis and Dr. Hibon", Journal of the Royal Statistical Society", Vol. 142, Part 2, pp. 127-129.

Saaty, T.L., 1980, The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation, McGraw-Hill, New York.

Schnaars, S.P., 1986, "An evaluation of rules of selecting on extrapolation model on yearly sales forecasts", Interfaces, Vol. 16, N° 6, pp. 100-107.

Williams, W.H. and M.L. Goodman, 1971, "A simple method for the construction of empirical confidence limits for economic forecasts", Journal of the American Statistical Association, 66, pp. 752-754.

Zarnowitz, V., 1984, "The accuracy of individual and group forecasts from business outlook surveys, Journal of Forecasting, 3, pp. 11-26.

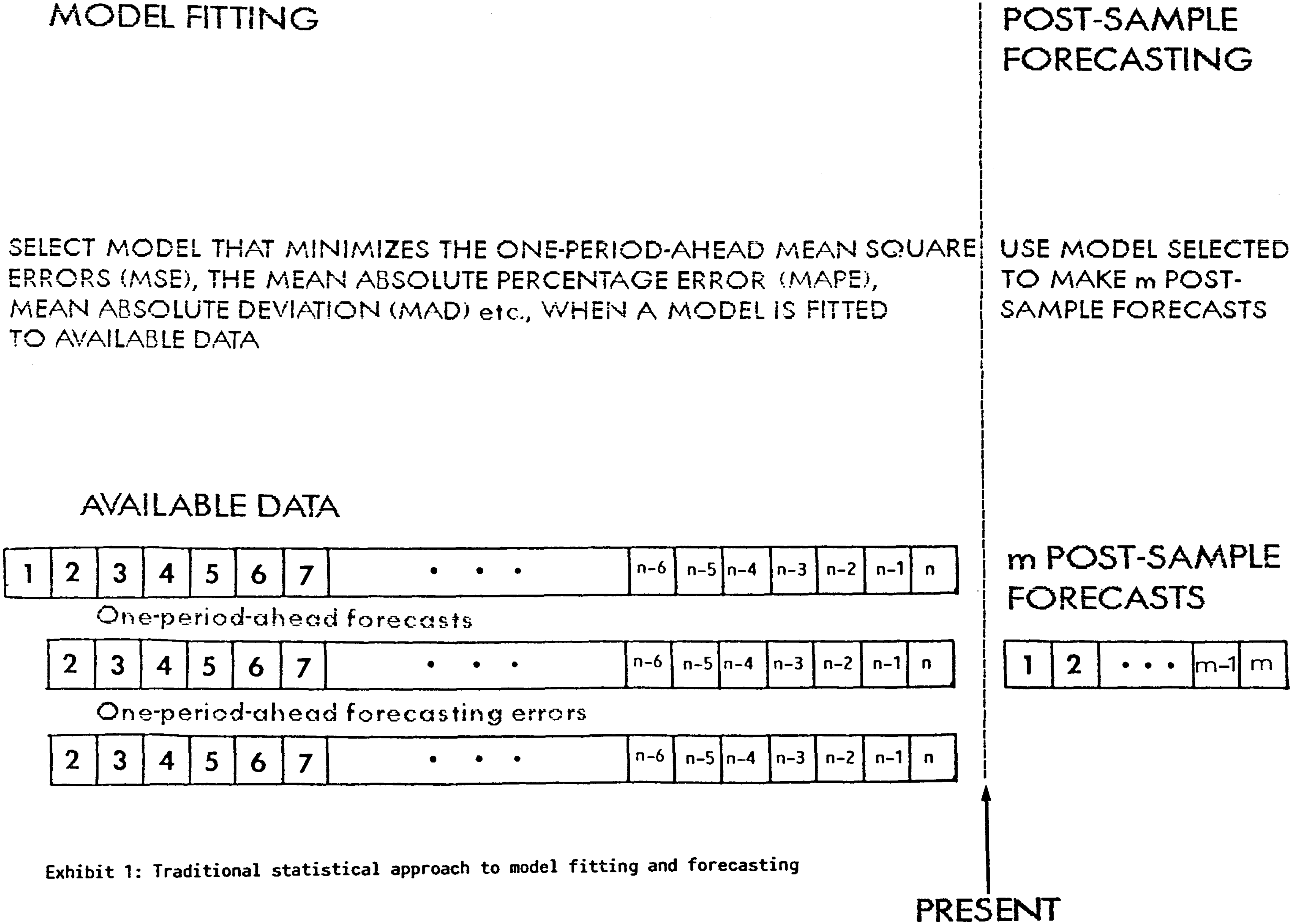
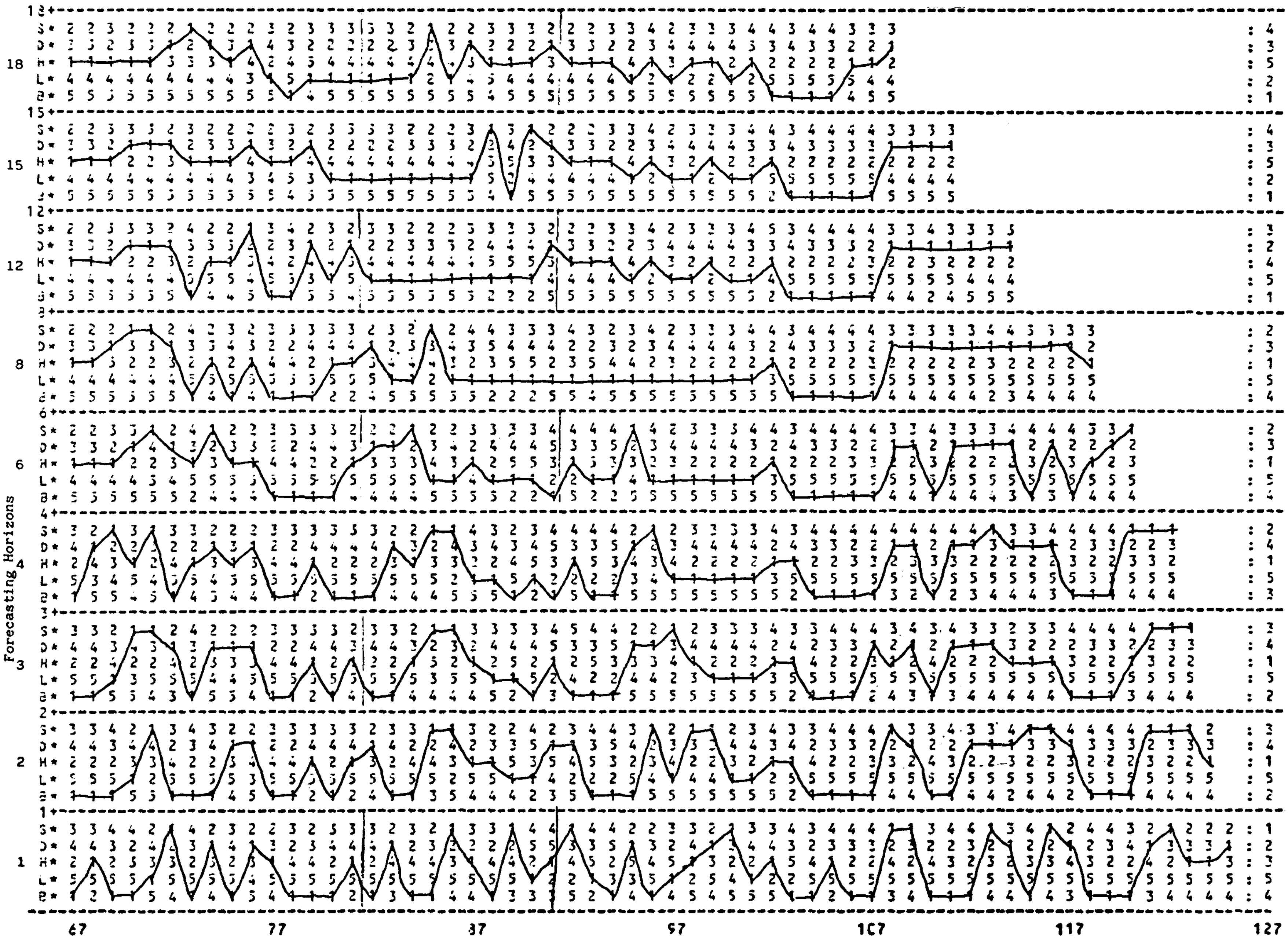


Exhibit 1: Traditional statistical approach to model fitting and forecasting

Exhibit 2: Ranking of five methods at different time periods and for several forecasting horizons



Serie : MNC44

S = Single Exponential Smoothing, D = Dampen Trend Exponential Smoothing, H = Holt's Exponential Smoothing, L = Linear Trend bases Regression, and B = Brown's Quadratic Exponential Smoothing

Exhibit 3: The international airline passengers 1949 - 1981

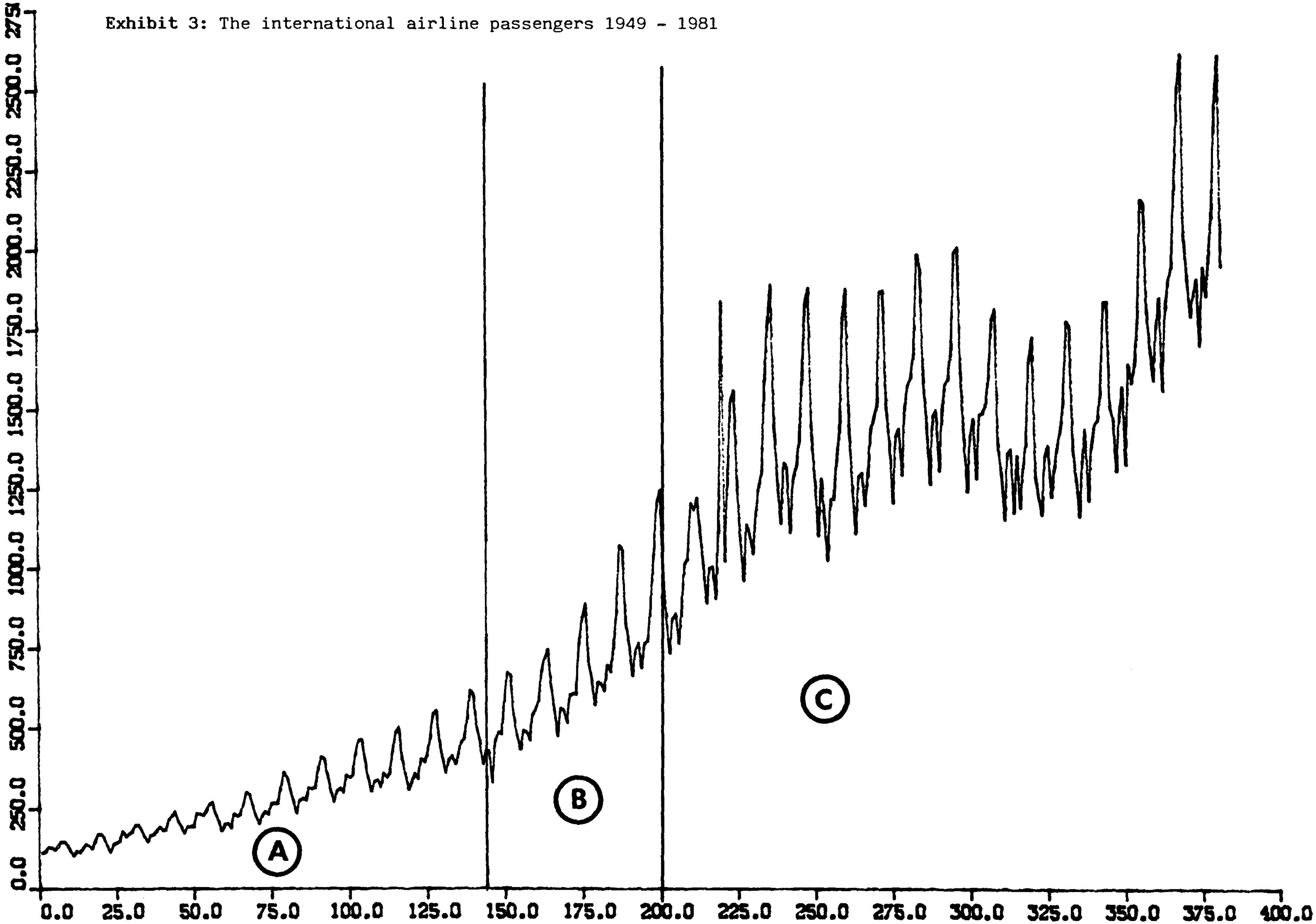


EXHIBIT 4:

Orders and Pulp Prices in Constant $

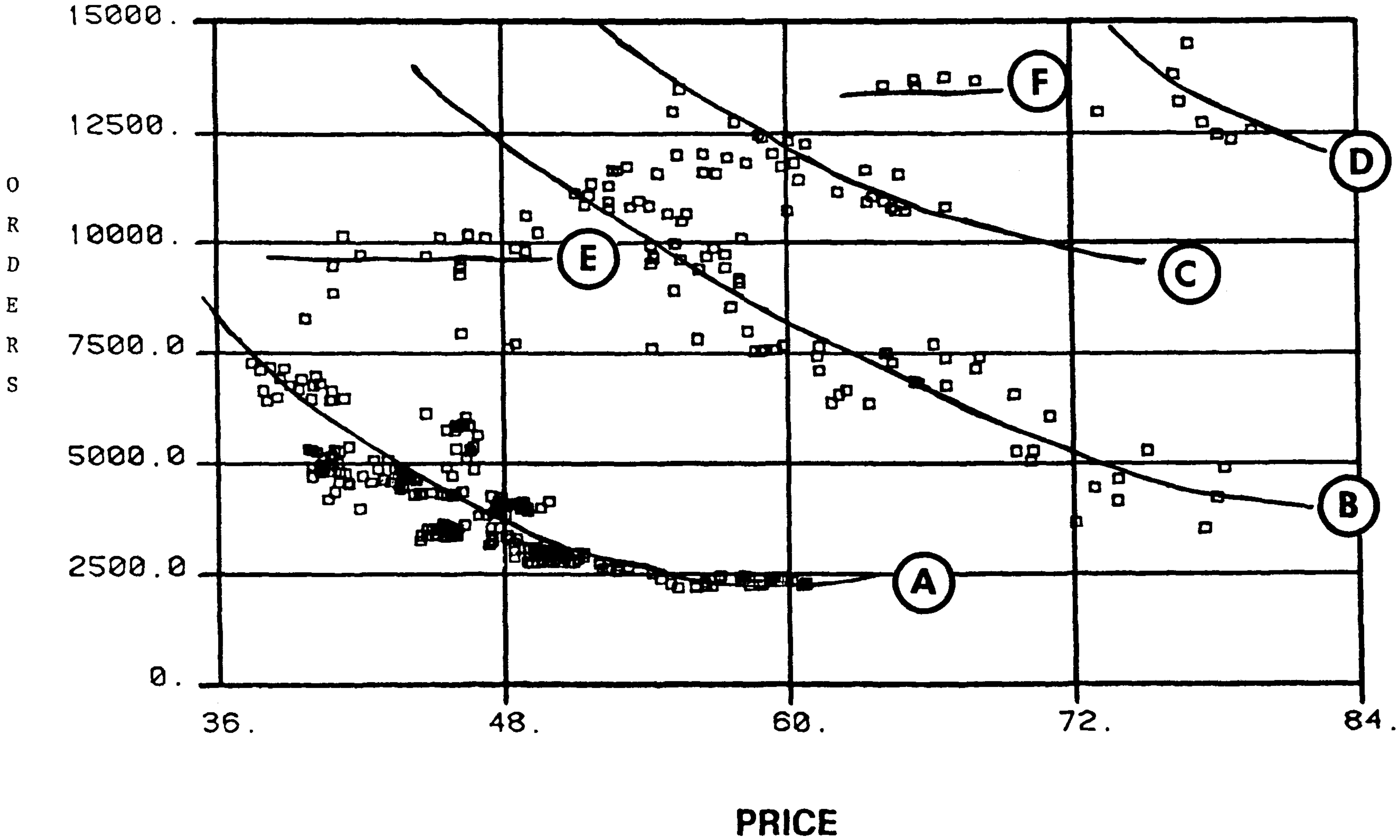
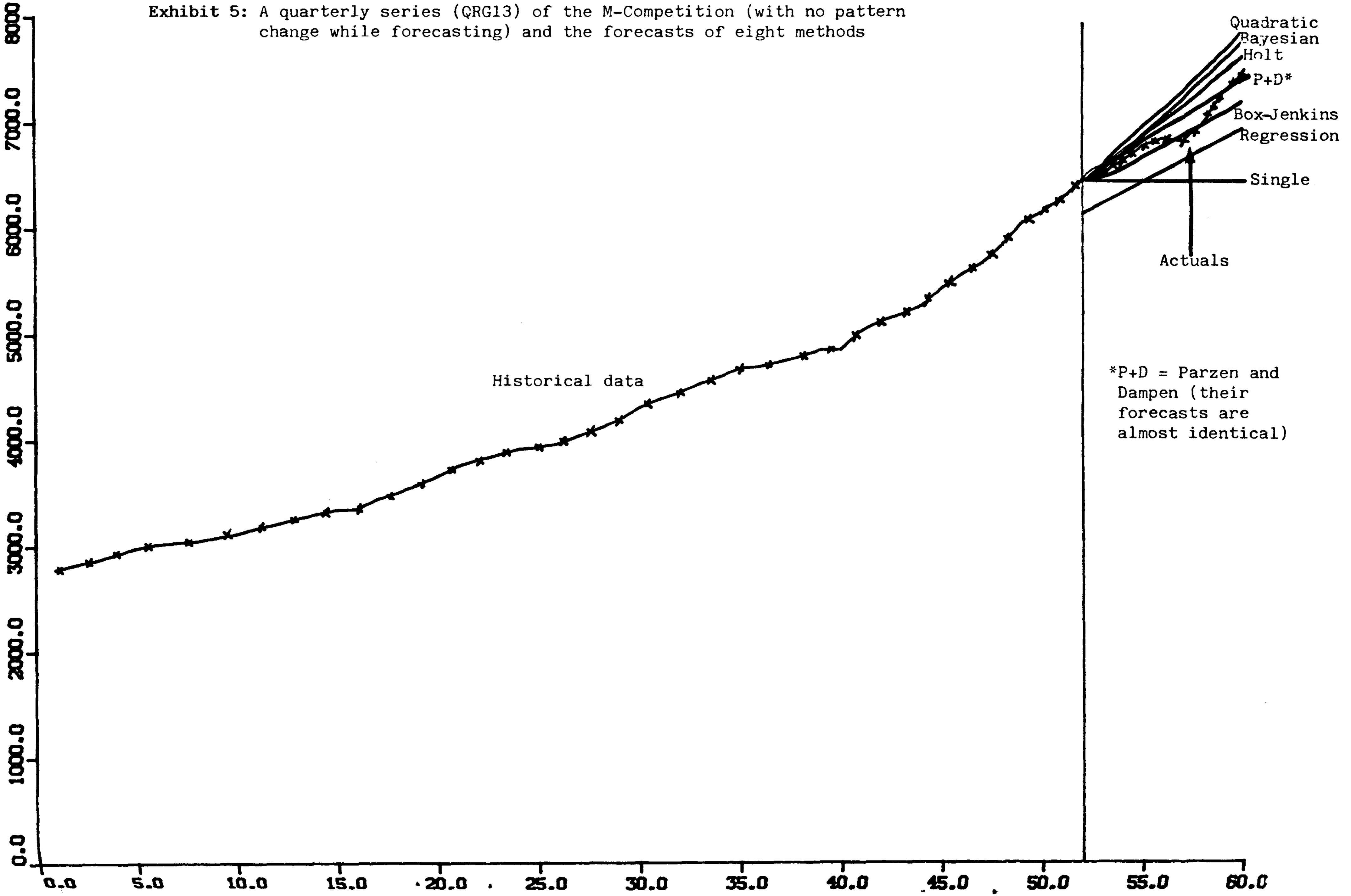


Exhibit 5: A quarterly series (QRG13) of the M-Competition (with no pattern change while forecasting) and the forecasts of eight methods



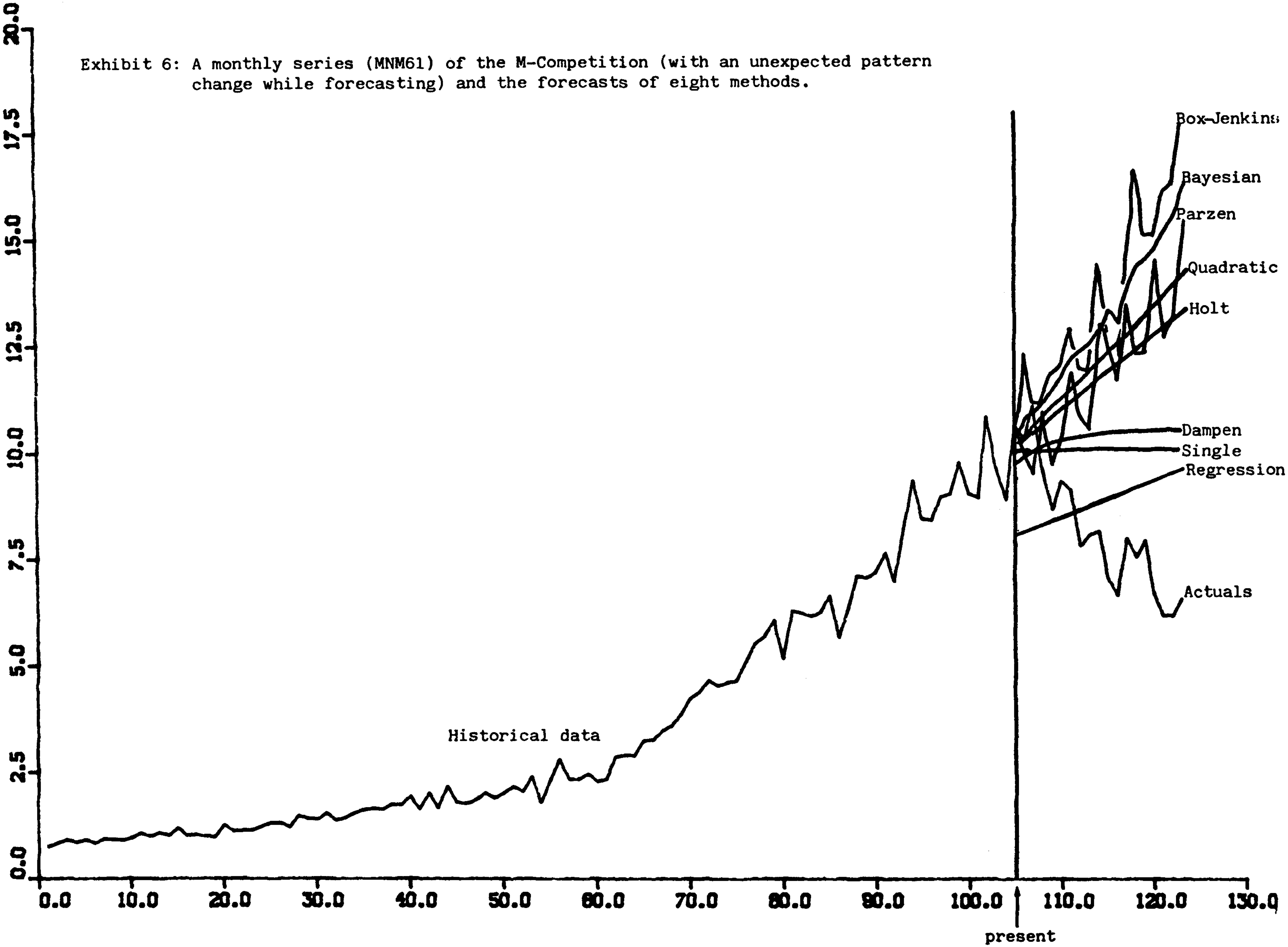


Exhibit 6: A monthly series (MNM61) of the M-Competition (with an unexpected pattern change while forecasting) and the forecasts of eight methods.

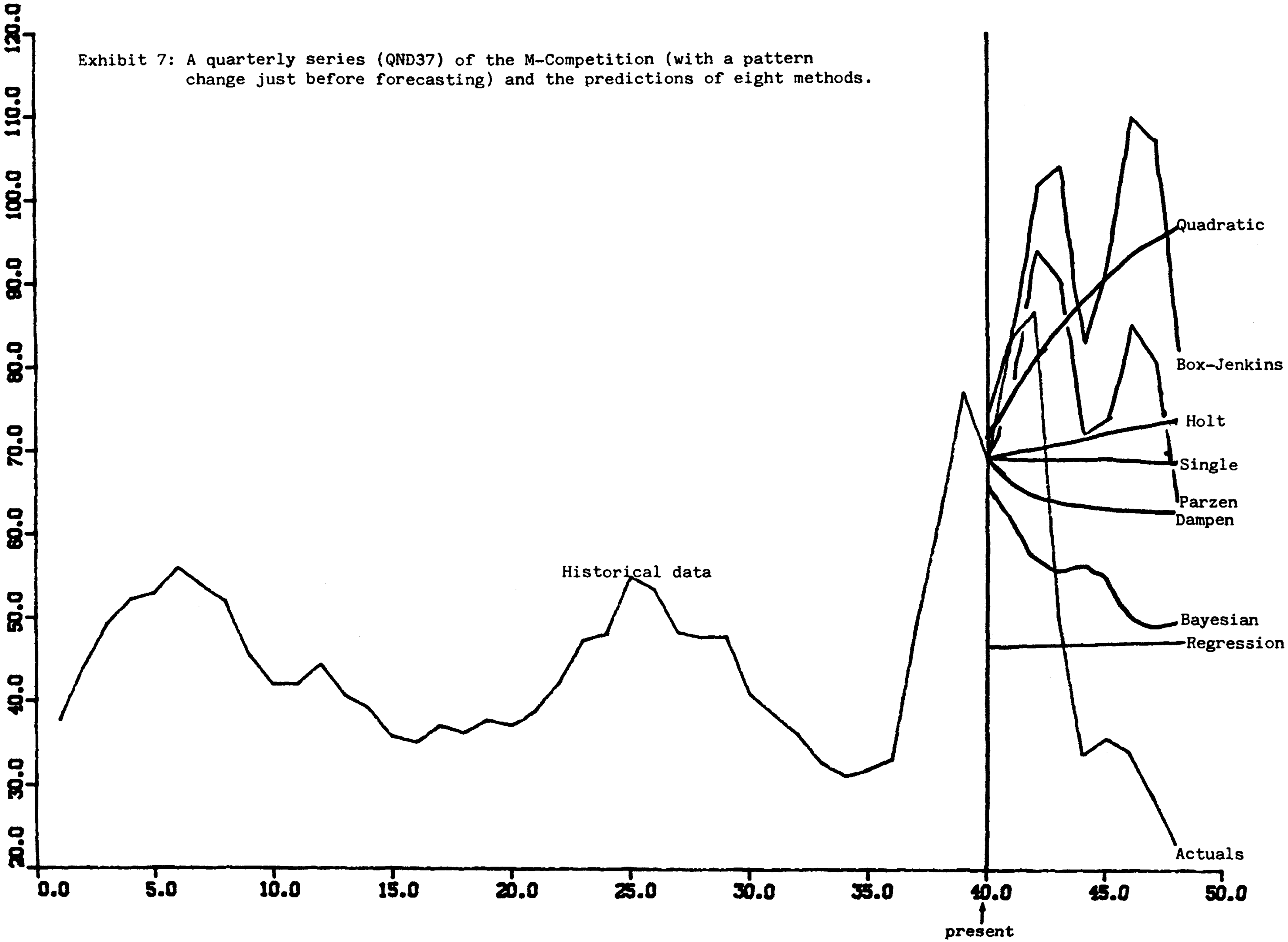


Exhibit 7: A quarterly series (QND37) of the M-Competition (with a pattern change just before forecasting) and the predictions of eight methods.

Exhibit 8 : A monthly series (MNC44) of the M-Competition (with a cyclical pattern change) and the forecasts of eight methods.

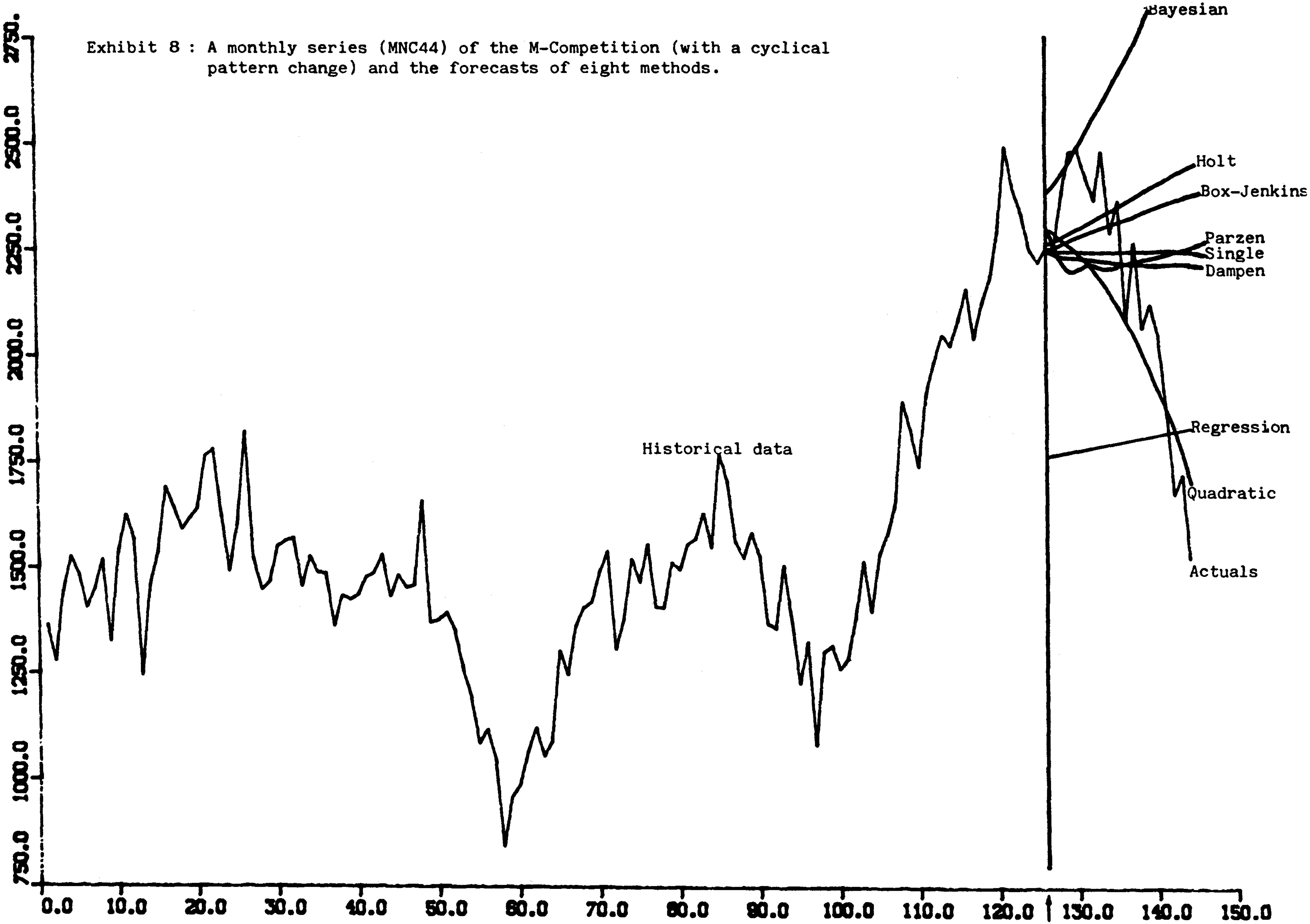


Exhibit 9: Major empirical evidence and its implications

Major findings	Empirical Evidence	Implications
1. Simple methods	Simple, automatic and inexpensive methods give realistic forecasts.	Use simple methods to a greater extent unless specific reasons that can be substantiated by concrete empirical evidence exists. For instance, use exponential smoothing methods.
2. Seasonality	Seasonality can be predicted accurately no matter what approach is being used.	Deseasonalize the data to develop a model and forecast. Then re-seasonalize forecasts.
3. Combining	Combining different methods (by a simple arithmetic average) improves forecasting accuracy and reduces the variance of errors.	No matter what the approach utilized use several methods and combine their forecasts. Choose methods in such a way as their forecasts will be as complementary (therefore independent) as possible.
4. Short versus Long term	Some models are more accurate for the short term (e.g., single exponential smoothing) others are more accurate for the long term (e.g., long memory ARARMA models).	In addition to traditional methods also use an AR(p) model where the length of p is large. Such AR(p) (called long memory) is appropriate for capturing and extrapolating the long term trend.
5. Dampening the trend	Dampening the trend improves forecasting accuracy.	Dampen the trend extrapolation using a dampen-trend exponential smoothing model.

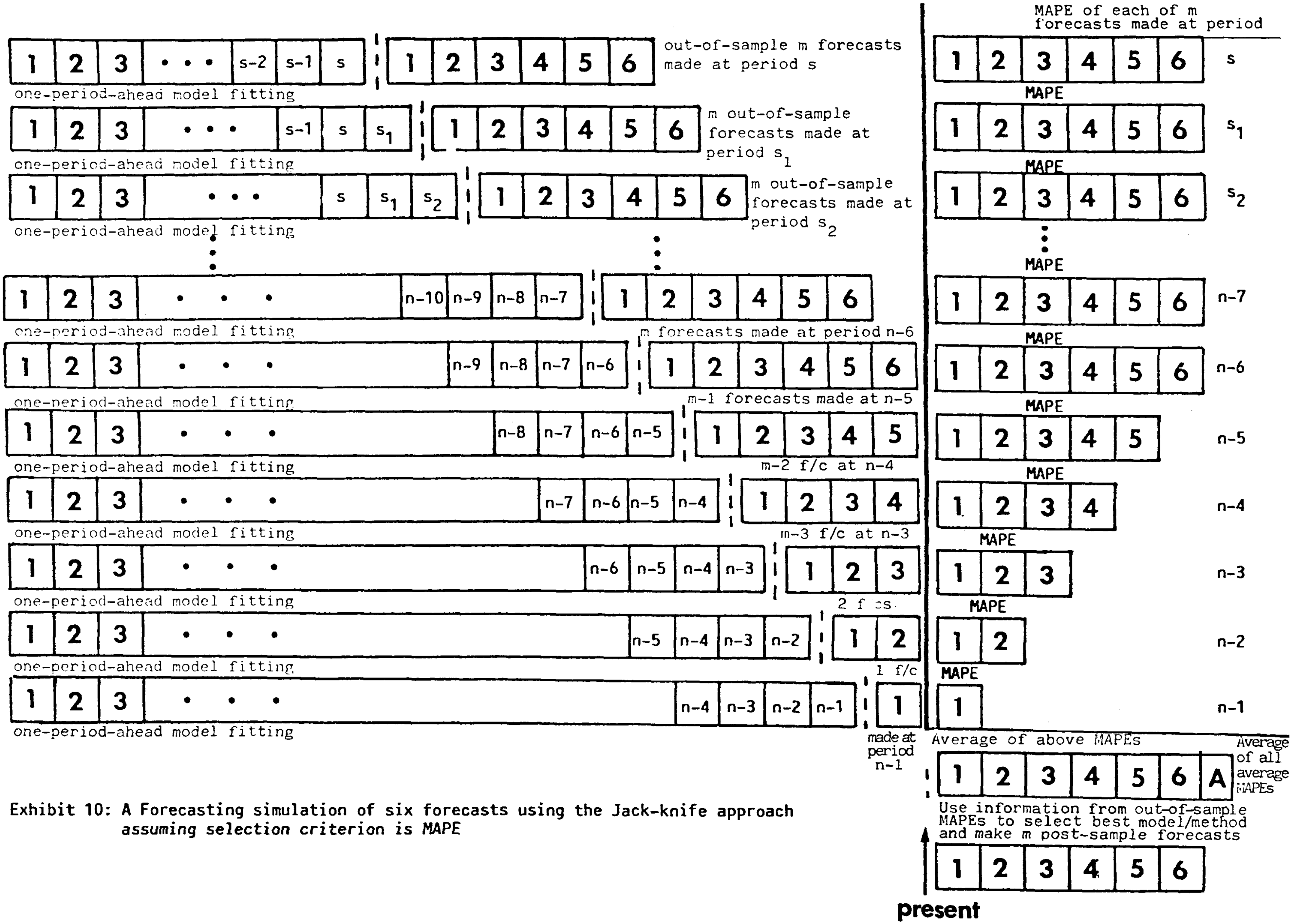


Exhibit 10: A Forecasting simulation of six forecasts using the Jack-knife approach
assuming selection criterion is MAPE

Exhibit 11: **PERFORMANCE OF FIVE METHODS AT THREE HORIZONS**

		1 Best	1 Worst	6 Best	6 Worst	18 Best	18 Worst	1-18 Average	MODEL FIT RMSE
R	S*	7	6	6	6	9	3		3.9
A	D*	8	4	14	3	9	3		8.5
N	H*	7	5	3	7	12	0		3.9
K	L*	8	18	8	6	12	1		8.2
S	B*	17	14	11	20	4	22		2.8* Smallest Root Mean Square Error (model fit)

Out-of-Sample MAPE

		1	6	18	1-18 Average	MAPE
M	S*	1.9	6.6	12.1	7.7	1.9
A	D*	1.7	5.8	12.7	7.4	2.8
P	H*	1.8	6.5	11.2	7.5	1.9
E	L*	2.1	6.4	15.4	8.6	2.5
	B*	1.7	6.0	26.9	11.5	1.5 * Small MAPE

Post-Sample MAPE

		1	6	18	1-18 Average
R					
E	S*	5.1	13.2	16.8	11.2
A	D*	0.2	6.1	32.5	16.7
L	H*	4.9	12.9	17.2	11.2
	L*	2.5	2.6	44.4	22.8
M	B*	3.7	54.3	220.7	104.9
A	C*	0.5	21.0	22.1	10.9
P	M*	DBH	DBL	HSD	
E					

S = Single, D = Dampen-trend, H = Holt, L = Long-term trend, B = Brown's Quadratic
M = Methods selected for post-sample forecasting

AVERAGE MAPE: ALL DATA (111)

METHODS	FITTING MODEL	Forecasting Horizons 1	2	3	4	5	6	8	12	15	18	Average of Forecasting Horizons 1-4	1-6	1-8	1-12	1-15	1-18	n(max)
Single exponential smoothing	8.6	7.8	10.8	13.1	14.5	15.7	17.2	16.5	13.6	29.3	30.1	11.6	13.2	14.1	14.0	15.3	16.8	111
Dampen-trend exponential smoot.	10.1	7.8	10.2	12.4	14.4	15.9	16.8	18.1	14.0	30.6	30.6	11.2	12.9	14.2	14.3	15.7	17.2	111
Holt's linear exponential sm.	8.6	7.9	10.5	13.2	15.1	17.3	19.0	23.1	16.5	35.6	35.2	11.7	13.8	16.1	16.4	18.0	19.7	111
Long term memory AR(p) model	6.8	9.6	8.6	10.3	12.2	13.6	14.1	14.7	14.7	18.0	24.5	10.2	11.4	12.3	12.7	13.3	14.3	111
Brown's quadratic expon. smoot.	8.7	8.8	11.8	15.0	16.9	21.9	24.1	35.7	29.7	56.1	63.6	13.1	16.4	20.3	22.2	25.9	30.2	111
Above five methods combined	8.8	7.8	9.7	11.0	12.9	14.7	15.6	17.6	14.0	30.1	32.1	10.3	11.9	13.4	13.8	15.3	17.0	111
Automatic A.E.P. filter	10.8	9.8	11.3	13.7	15.1	16.9	18.8	23.3	16.2	30.2	33.9	12.5	14.3	16.3	16.2	17.4	19.0	111
Bayesian forecasting	13.3	10.3	12.8	13.6	14.4	16.2	17.1	19.2	16.1	27.5	30.6	12.8	14.1	15.2	15.0	16.1	17.6	111
Box-Jenkins'ARIMA models	0.0	10.3	10.7	11.4	14.5	16.4	17.1	18.9	16.4	26.2	34.2	11.7	13.4	14.8	15.1	16.3	18.0	111
Lewandowski's FORSYS	12.3	11.6	12.8	14.5	15.3	16.6	17.6	18.9	17.0	33.0	28.6	13.5	14.7	15.5	15.6	17.2	18.6	111
Parzen's ARARMA models	8.9	10.6	10.7	10.7	13.5	14.3	14.7	16.0	13.7	22.5	26.5	11.4	12.4	13.3	13.4	14.3	15.4	111
Method with best MSE Model fit	6.7	8.4	8.3	11.2	13.8	14.3	16.0	17.8	17.0	33.9	34.6	10.4	12.0	13.5	14.1	15.9	18.0	111
Method with best MAPE Model fit	6.1	8.4	8.9	11.9	15.0	15.1	16.7	19.5	15.4	31.6	31.6	11.0	12.7	14.3	14.4	15.8	17.5	111
Best MSE ALL F/Cs out-of-sample	8.4	7.9	9.0	12.4	15.2	17.0	17.7	25.3	14.2	20.7	25.8	11.1	13.2	15.7	15.1	15.5	16.4	111
Best MAPE ALL F/Cs O-of-S	8.6	7.9	9.0	12.3	15.4	17.0	17.8	25.1	14.3	20.7	25.6	11.1	13.2	15.7	15.1	15.5	16.4	111
Best MSE each F/C horiz. O-of-S	8.5	7.8	8.7	10.5	12.9	13.7	13.5	15.6	14.4	20.1	20.7	10.0	11.2	12.3	12.6	13.3	13.9	111
Best MAPE each F/C hor. O-of-S	8.6	7.7	8.7	10.6	12.9	13.8	13.5	15.9	14.3	20.2	20.5	10.0	11.1	12.2	12.5	13.2	13.8	111
Best Median each F/C hor.O-of-S	8.6	8.3	9.3	10.8	13.7	14.2	14.9	16.2	14.8	21.3	26.1	10.5	11.9	13.0	13.2	13.9	15.1	111
Combine 2Methods with best MAPE	8.7	7.7	9.0	10.7	12.8	13.1	13.5	15.5	13.1	19.8	20.6	10.1	11.1	12.3	12.5	13.1	13.7	111
Combine 3Methods with best MAPE	8.8	7.9	9.1	10.0	12.4	12.7	13.5	15.7	12.6	19.0	20.7	9.8	10.9	12.2	12.4	13.0	13.6	111
Combine 4Methods with best MAPE	8.8	7.8	8.6	10.2	12.4	12.8	13.7	16.2	12.9	19.3	20.8	9.7	10.9	12.3	12.5	13.2	13.8	111
Comb. meth. w. best MSEMAPERank	8.6	7.7	8.7	10.0	12.7	13.0	13.8	15.3	14.4	20.3	20.7	9.8	11.0	12.2	12.5	13.2	13.7	111
PRIOR unless you can REJECT Ho	8.3	7.1	8.3	10.3	12.9	13.4	13.6	15.0	14.7	20.9	20.8	9.7	10.9	12.1	12.4	13.2	13.8	111
Methods within confidence Inter	8.5	7.4	8.3	10.2	12.3	12.9	14.0	15.3	14.0	20.0	20.4	9.6	10.9	12.2	12.4	13.1	13.7	111

Exhibit 13: Within method model selection using as criterion the model that minimizes the MSE for each forecasting horizon

A V E R A G E M A P E : (ALL DATA)

METHODS	MODEL FITTING	Forecasting Horizons 1	2	3	4	5	6	8	12	15	18	Average of Forecasting Horizons 1-4	1-6	1-8	1-12	1-15	1-18	n(max)
PARZEN ARARMA (M-Competition)	8.9	10.6	10.7	10.7	13.5	14.3	14.7	16.0	13.7	22.5	26.5	11.4	12.4	13.3	13.4	14.3	15.4	111
SINGLE (Optimal Model Fitting)	8.6	7.8	10.8	13.1	14.5	15.7	17.2	16.5	13.6	29.3	30.1	11.6	13.2	14.1	14.0	15.3	16.8	111
SINGLE (Optimal Out-of-sample)	8.6	8.0	10.7	12.6	13.7	14.7	16.1	15.1	13.6	25.1	25.3	11.3	12.6	13.4	13.3	14.3	15.5	111
DAMPEN-Trend(Opt.Model Fitting)	10.1	7.8	10.2	12.4	14.4	15.9	16.8	18.1	14.0	30.6	30.6	11.2	12.9	14.2	14.3	15.7	17.2	111
DAMPEN-Trend(Opt.Out-of-sample)	10.1	8.2	9.6	12.1	12.7	14.8	17.2	18.6	15.5	27.3	27.3	10.6	12.4	13.9	14.2	15.2	16.6	111
HOLT (Optimal Model Fitting)	8.6	7.9	10.5	13.2	15.1	17.3	19.0	23.1	16.5	35.6	35.2	11.7	13.8	16.1	16.4	18.0	19.7	111
HOLT (Optimal Out-of-sample)	8.6	7.9	10.3	11.9	13.6	15.0	15.7	17.8	14.0	28.0	26.8	10.9	12.4	13.8	14.1	15.3	16.4	111
QUADRATIC (Optim.Model Fitting)	8.7	8.8	11.8	15.0	16.9	21.9	24.1	35.7	29.7	56.1	63.6	13.1	16.4	20.3	22.2	25.9	30.2	111
QUADRATIC (Optim.Out-of-sample)	8.7	8.8	11.0	14.9	14.6	16.8	18.6	22.3	26.5	36.2	47.2	12.3	14.1	15.8	16.9	19.2	22.3	111

A V E R A G E M A P E : YEARLY DATA (20)

METHODS	MODEL FITTING	Forecasting Horizons 1	2	3	4	5	6	8	12	15	18	Average of Forecasting Horizons 1-4	1-6	1-8	1-12	1-15	1-18	n(max)
PARZEN ARARMA (M-Competition)	9.6	7.6	7.7	12.8	16.0	20.5	18.0	0.0	0.0	0.0	0.0	11.0	13.8	13.8	13.8	13.8	13.8	20
SINGLE (Optimal Model Fitting)	11.4	6.2	9.1	16.3	21.0	23.6	25.4	0.0	0.0	0.0	0.0	13.1	16.9	16.9	16.9	16.9	16.9	20
SINGLE (Optimal Out-of-sample)	25.2	6.5	8.5	16.3	21.1	24.0	24.8	0.0	0.0	0.0	0.0	13.1	16.8	16.8	16.8	16.8	16.8	20
DAMPEN-Trend(Opt.Model Fitting)	15.1	6.9	9.6	15.2	20.3	23.4	20.9	0.0	0.0	0.0	0.0	13.0	16.0	16.0	16.0	16.0	16.0	20
DAMPEN-Trend(Opt.Out-of-sample)	12.6	6.6	7.1	11.9	18.7	26.1	24.4	0.0	0.0	0.0	0.0	11.1	15.8	15.8	15.8	15.8	15.8	20
HOLT (Optimal Model Fitting)	12.9	5.6	7.2	11.9	16.2	19.0	16.5	0.0	0.0	0.0	0.0	10.2	12.7	12.7	12.7	12.7	12.7	20
HOLT (Optimal Out-of-sample)	11.3	5.7	6.7	10.9	14.1	17.7	15.9	0.0	0.0	0.0	0.0	9.4	11.8	11.8	11.8	11.8	11.8	20
QUADRATIC (Optim.Model Fitting)	11.1	7.0	8.6	11.8	16.0	20.7	17.4	0.0	0.0	0.0	0.0	10.9	13.6	13.6	13.6	13.6	13.6	20
QUADRATIC (Optim.Out-of-sample)	10.9	7.3	6.8	11.3	12.4	15.5	15.1	0.0	0.0	0.0	0.0	9.5	11.4	11.4	11.4	11.4	11.4	20

A V E R A G E M A P E : QUARTERLY DATA (23)

METHODS	MODEL FITTING	Forecasting Horizons 1	2	3	4	5	6	8	12	15	18	Average of Forecasting Horizons 1-4	1-6	1-8	1-12	1-15	1-18	n(max)
PARZEN ARARMA (M-Competition)	7.7	6.8	7.6	12.0	16.5	21.1	20.4	21.0	0.0	0.0	0.0	10.7	14.1	16.7	16.7	16.7	16.7	23
SINGLE (Optimal Model Fitting)	7.7	9.0	12.0	14.4	20.5	21.0	21.9	22.6	0.0	0.0	0.0	14.0	16.5	18.5	18.5	18.5	18.5	23
SINGLE (Optimal Out-of-sample)	9.5	9.9	11.6	12.9	16.7	18.0	18.1	18.3	0.0	0.0	0.0	12.8	14.5	16.2	16.2	16.2	16.2	23
DAMPEN-Trend(Opt.Model Fitting)	9.6	8.8	8.6	11.9	19.7	22.3	24.8	26.6	0.0	0.0	0.0	12.2	16.0	19.3	19.3	19.3	19.3	23
DAMPEN-Trend(Opt.Out-of-sample)	8.3	9.5	8.6	12.7	13.7	17.1	21.9	22.0	0.0	0.0	0.0	11.1	13.9	16.6	16.6	16.6	16.6	23
HOLT (Optimal Model Fitting)	7.2	9.2	10.4	17.1	25.1	30.3	32.2	39.2	0.0	0.0	0.0	15.4	20.7	25.9	25.9	25.9	25.9	23
HOLT (Optimal Out-of-sample)	9.0	7.4	10.6	14.4	20.3	21.1	21.5	21.2	0.0	0.0	0.0	13.2	15.9	17.9	17.9	17.9	17.9	23
QUADRATIC (Optim.Model Fitting)	7.9	11.1	12.5	21.1	32.0	39.2	46.0	66.6	0.0	0.0	0.0	19.2	27.0	35.6	35.6	35.6	35.6	23
QUADRATIC (Optim.Out-of-sample)	7.3	10.0	12.1	22.4	25.2	28.1	29.5	34.0	0.0	0.0	0.0	17.4	21.2	24.6	24.6	24.6	24.6	23

A V E R A G E M A P E : MONTHLY DATA (68)

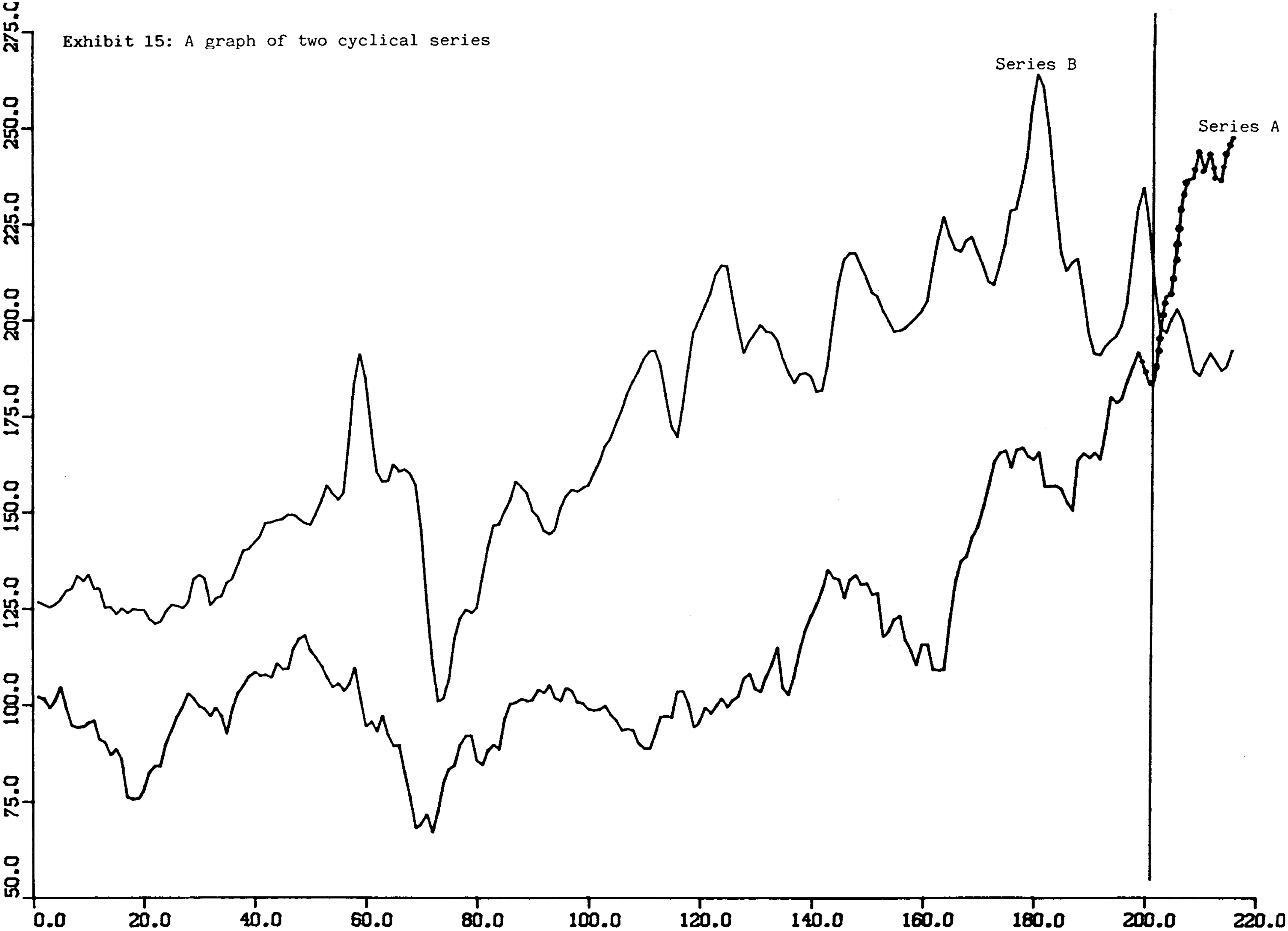
METHODS	MODEL FITTING	Forecasting Horizons 1	2	3	4	5	6	8	12	15	18	Average of Forecasting Horizons 1-4	1-6	1-8	1-12	1-15	1-18	n(max)
PARZEN ARARMA (M-Competition)	9.0	12.7	12.6	9.6	11.7	10.2	11.8	14.3	13.7	22.5	26.5	11.7	11.4	12.1	12.6	13.9	15.4	68
SINGLE (Optimal Model Fitting)	8.0	7.9	10.9	11.7	10.6	11.6	13.2	14.4	13.6	29.3	30.1	10.3	11.0	12.0	12.6	14.5	16.5	68
SINGLE (Optimal Out-of-sample)	8.5	7.8	11.1	11.5	10.5	10.8	12.8	14.0	13.6	25.1	25.3	10.2	10.8	11.6	12.1	13.7	15.2	68
DAMPEN-Trend(Opt.Model Fitting)	8.7	7.8	11.0	11.8	10.9	11.6	12.9	15.2	14.0	30.6	30.6	10.4	11.0	12.1	13.0	15.0	17.1	68
DAMPEN-Trend(Opt.Out-of-sample)	8.0	8.2	10.7	12.0	10.5	10.7	13.5	17.4	15.5	27.3	27.3	10.4	10.9	12.5	13.4	14.9	16.7	68
HOLT (Optimal Model Fitting)	7.9	8.2	11.5	12.3	11.4	12.5	15.2	17.7	16.5	35.6	35.2	10.9	11.8	13.5	14.8	17.2	19.5	68
HOLT (Optimal Out-of-sample)	8.7	8.7	11.3	11.3	11.2	12.2	13.7	16.7	14.0	28.0	26.8	10.6	11.4	12.8	13.6	15.3	16.6	68
QUADRATIC (Optim.Model Fitting)	8.2	8.6	12.5	13.8	12.1	16.3	18.7	25.3	29.7	56.1	63.6	11.7	13.7	16.6	20.4	25.7	31.0	68
QUADRATIC (Optim.Out-of-sample)	8.4	8.8	11.8	13.5	11.6	13.4	16.0	18.4	26.5	36.2	47.2	11.4	12.5	13.9	16.0	19.2	23.0	68

Exhibit 14: Within-method model selection (the "best" out-of-sample model for each forecasting horizon) and among methods "best" model selection (the "best" model selection (the "best" three methods)

AVERAGE MAPE: ALL DATA (111)

METHODS	MODEL FITTING	Forecasting Horizons 1	2	3	4	5	6	8	12	15	18	Average of Forecasting Horizons 1-4	1-6	1-8	1-12	1-15	1-18	n(max)
Combine 3Methods with best MAPE	8.6	7.9	9.1	10.0	12.4	12.7	13.5	15.7	12.6	19.0	20.7	9.8	10.9	12.2	12.4	13.0	13.6	111
Comb. 3Meth. Single within-Mod.	8.6	7.6	8.7	10.0	12.1	12.6	13.3	15.5	12.3	18.8	20.2	9.6	10.7	12.0	12.1	12.5	13.1	111
Comb. 3Meth. Sing+Holt within-M	8.6	7.5	8.9	10.0	12.1	12.2	13.1	15.8	12.5	18.8	20.0	9.6	10.6	11.9	11.9	12.4	13.0	111
C. 3Meth. Sing+Holt+Brown w-M.	8.6	7.5	8.7	10.0	12.0	12.2	13.3	15.7	12.6	18.6	19.6	9.5	10.6	11.9	11.8	12.4	13.0	111
C. 3Met. Sing+Holt+Brown+Dampen	8.6	7.4	8.7	10.0	11.8	12.0	12.7	15.4	12.4	18.6	19.5	9.4	10.4	11.7	11.6	12.2	12.8	111

Exhibit 15: A graph of two cyclical series



1984

84/01	Arnoud DE MEYER	"A technological life-cycle to the organisational factors determining gatekeeper activities" , November 1983.
84/02	Jeffrey SACHS and Charles A. WYPLOSZ	"La politique budgétaire et le taux de change réel", November 1983.
84/03	Jeffrey SACHS and Charles A. WYPLOSZ	"Real exchange rate effects of fiscal policy", December 1983.
84/04	Gabriel A. HAWAWINI	"European equity markets: a review of the evidence on price behavior and efficiency", February 1984.
84/05	Charles A. WYPLOSZ	"Capital controls and balance of payments crises", February 1984.
84/06	Gabriel A. HAWAWINI	"An uncertainty model of the professional partnership", November 1983.
84/07	Gabriel A. HAWAWINI	"The geometry of risk aversion", October 1983.
84/08	Gabriel A. HAWAWINI, Pierre MICHEL and Claude J. VIALLET	"Risk, Return and equilibrium of the NYSE: update, robustness of results and extensions" December 1983.
84/09	Gabriel A. HAWAWINI, Claude J. VIALLET and Ashok VORA	"Industry influence on firm's investment in working capital: theory and evidence", January 1984.
84/10	Gabriel A. HAWAWINI and Pierre A. MICHEL	"Impact of the Belgian Financial Reporting Act of 1976 on the systematic risk of common stocks", January 1984.
84/11	Jean DERMINE	"On the measurement of the market value of a bank", April 1984.
84/12	Antonio M. BORGES	"Tax reform in Portugal: a general equilibrium analysis of the introduction of a value added tax", December 1984.
84/13	Arnoud DE MEYER and Kasra FERDOWS	"Integration of information systems in manufacturing", December 1984.

1985

85/01	Jean DERMINE	"The measurement of interest rate risk by financial intermediaries", December 1983, Revised December 1984.
85/02	Philippe A. NAERT and Els GIJSBRECHTS	"Diffusion model for new product introduction in existing markets" .
85/03	Philippe A. NAERT and Els GIJSBRECHTS	"Towards a decision support system for hierarchically allocating marketing resources across and within product groups" .
85/04	Philippe A. NAERT and Marcel WEVERBERGH	"Market share specification, estimation and validation: towards reconciling seemingly divergent views" .
85/05	Ahmet AYKAC, Marcel CORSTJENS, David GAUTSCHI and Ira HOROWITZ	"Estimation uncertainty and optimal advertising decisions", Second draft, April 1985.
85/06	Kasra FERDOWS	"The shifting paradigms of manufacturing: inventory, quality and now versatility", March 1985.
85/07	Kasra FERDOWS, Jeffrey G. MILLER, Jinchiro NAKANE and Thomas E.VOLLMANN.	"Evolving manufacturing strategies in Europe, Japan and North-America"
85/08	Spyros MAKRIDAKIS and Robert CARBONE	"Forecasting when pattern changes occur beyond the historical data" , April 1985.
85/09	Spyros MAKRIDAKIS and Robert CARBONE	"Sampling distribution of post-sample forecasting errors" , February 1985.
85/10	Jean DERMINE	"Portfolio optimization by financial intermediaries in an asset pricing model".
85/11	Antonio M. BORGES and Alfredo M. PEREIRA	"Energy demand in Portuguese manufacturing: a two-stage model".
85/12	Arnoud DE MEYER	"Defining a manufacturing strategy - a survey of European manufacturers".
85/13	Arnoud DE MEYER	"Large European manufacturers and the management of R & D".
85/14	Ahmet AYKAC, Marcel CORSTJENS, David GAUTSCHI and Douglas L. MacLACHLAN	"The advertising-sales relationship in the U.S. cigarette industry: a comparison of correlational and causality testing approaches".
85/15	Arnoud DE MEYER and Roland VAN DIERDONCK	"Organizing a technology jump or overcoming the technological hurdle".
85/16	Herwig M. LANGOHR and Antony M. SANTOMERO	"Commercial bank refinancing and economic stability: an analysis of European features".
85/17	Manfred F.R. KETS DE VRIES and Danny MILLER	"Personality, culture and organization".
85/18	Manfred F.R. KETS DE VRIES	"The darker side of entrepreneurship".
85/19	Manfred F.R. KETS DE VRIES and Dany MILLER	"Narcissism and leadership: an object relations perspective".
85/20	Manfred F.R. KETS DE VRIES and Dany MILLER	"Interpreting organizational texts".

85/21	Herwig M. LANGOHR and Claude J. VIALLET	**"Nationalization, compensation and wealth transfers: France 1981-1982"** 1, Final version July 1985.
85/22	Herwig M. LANGOHR and B. Espen ECKBO	**"Takeover premiums, disclosure regulations, and the market for corporate control. A comparative analysis of public tender offers, controlling-block trades and minority buyout in France"**, July 1985.
85/23	Manfred F.R. KETS DE VRIES and Dany MILLER	**"Barriers to adaptation: personal, cultural and organizational perspectives".**
85/24	Spyros MAKRIDAKIS	**"The art and science of forecasting: an assessment and future directions".**
85/25	Gabriel HAWAWINI	**"Financial innovation and recent developments in the French capital markets"**, October 1985.
85/26	Karel O. COOL and Dan E. SCHENDEL	**"Patterns of competition, strategic group formation and the performance case of the US pharmaceutical industry, 1963-1982"**, October 1985.
85/27	Arnoud DE MEYER	**"European manufacturing: a comparative study (1985)".**

1986

86/01	Arnoud DE MEYER	**"The R & D/Production interface".**
86/02	Philippe A. NAERT Marcel WEVERBERGH and Guido VERSWIJVEL	**"Subjective estimation in integrating communication budget and allocation decisions: a case study"**, January 1986.
86/03	Michael BRIMM	**"Sponsorship and the diffusion of organizational innovation: a preliminary view".**
86/04	Spyros MAKRIDAKIS and Michèle HIBON	**"Confidence intervals: an empirical investigation for the series in the M-Competition" .**
86/05	Charles A. WYPLOSZ	**"A note on the reduction of the workweek"**, July 1985.
86/06	Francesco GIAVAZZI, Jeff R. SHEEN and Charles A. WYPLOSZ	**"The real exchange rate and the fiscal aspects of a natural resource discovery"**, Revised version: February 1986.
86/07	Douglas L. MacLACHLAN and Spyros MAKRIDAKIS	**"Judgmental biases in sales forecasting"**, February 1986.
86/08	José de la TORRE and David H. NECKAR	**"Forecasting political risks for international operations"**, Second Draft: March 3, 1986.
86/09	Philippe C. HASPESLAGH	**"Conceptualizing the strategic process in diversified firms: the role and nature of the corporate influence process"**, February 1986.
86/10	R. MOENART, Arnoud DE MEYER, J. BARBE and D. DESCHOOLMEESTER.	**"Analysing the issues concerning technological de-maturity".**
86/11	Philippe A. NAERT and Alain BULTEZ	**"From "Lydiametry" to "Pinkhamization": misspecifying advertising dynamics rarely affects profitability".**
86/12	Roger BETANCOURT and David GAUTSCHI	**"The economics of retail firms"**, Revised April 1986.
86/13	S.P. ANDERSON and Damien J. NEVEN	**"Spatial competition à la Cournot".**
86/14	Charles WALDMAN	**"Comparaison internationale des marges brutes du commerce"**, June 1985.
86/15	Mihkel TOMBAK and Arnoud DE MEYER	**"How the managerial attitudes of firms with FMS differ from other manufacturing firms: survey results"**, June 1986.
86/16	B. Espen ECKBO and Herwig M. LANGOHR	**"Les primes des offres publiques, la note d'information et le marché des transferts de contrôle des sociétés".**
86/17	David B. JEMISON	**"Strategic capability transfer in acquisition integration"**, May 1986.
86/18	James TEBOUL and V. MALLERET	**"Towards an operational definition of services"**, 1986.
86/19	Rob R. WEITZ	**"Nostradamus: a knowledge-based forecasting advisor".**
86/20	Albert CORHAY, Gabriel HAWAWINI and Pierre A. MICHEL	**"The pricing of equity on the London stock exchange: seasonality and size premium"**, June 1986.
86/21	Albert CORHAY, Gabriel A. HAWAWINI and Pierre A. MICHEL	**"Risk-premia seasonality in U.S. and European equity markets"**, February 1986.
86/22	Albert CORHAY, Gabriel A. HAWAWINI and Pierre A. MICHEL	**"Seasonality in the risk-return relationships some international evidence"**, July 1986.
86/23	Arnoud DE MEYER	**"An exploratory study on the integration of information systems in manufacturing"**, July 1986.
86/24	David GAUTSCHI and Vithala R. RAO	**"A methodology for specification and aggregation in product concept testing"**, July 1986.
86/25	H. Peter GRAY and Ingo WALTER	**"Protection"**, August 1986.
86/26	Barry EICHENGREEN and Charles WYPLOSZ	**"The economic consequences of the Franc Poincare"**, September 1986.

86/27	Karel COOL and Ingemar DIERICKX	"Negative risk-return relationships in business strategy: paradox or truism?", October 1986.
86/28	Manfred KETS DE VRIES and Danny MILLER	"Interpreting organizational texts.
86/29	Manfred KETS DE VRIES	"Why follow the leader?".
86/30	Manfred KETS DE VRIES	"The succession game: the real story.
86/31	Arnoud DE MEYER	"Flexibility: the next competitive battle".
86/32	Karel COOL and Dan SCHENDEL	Performance differences among strategic group members", October 1986.
86/33	Ernst BALTENSPERGER and Jean DERMINE	"The role of public policy in insuring financial stability: a cross-country, comparative perspective", August 1986, Revised November 1986.
86/34	Philippe HASPESLAGH and David JEMISON	"Acquisitions: myths and reality", July 1986.
86/35	Jean DERMINE	"Measuring the market value of a bank, a primer", November 1986.
86/36	Albert CORHAY and Gabriel HAWAWINI	"Seasonality in the risk-return relationship: some international evidence", July 1986.
86/37	David GAUTSCHI and Roger BETANCOURT	"The evolution of retailing: a suggested economic interpretation".
86/38	Gabriel HAWAWINI	"Financial innovation and recent developments in the French capital markets", Updated: September 1986.
86/39	Gabriel HAWAWINI Pierre MICHEL and Albert CORHAY	"The pricing of common stocks on the Brussels stock exchange: a re-examination of the evidence", November 1986.
86/40	Charles WYPLOSZ	"Capital flows liberalization and the EMS, a French perspective", December 1986.
86/41	Kasra FERDOWS and Wickham SKINNER	"Manufacturing in a new perspective", July 1986.
86/42	Kasra FERDOWS and Per LINDBERG	"FMS as indicator of manufacturing strategy", December 1986.
86/43	Damien NEVEN	"On the existence of equilibrium in hotelling's model", November 1986.

1987

87/01	Manfred KETS DE VRIES	"Prisoners of leadership".
87/02	Claude VIALLET	"An empirical investigation of international asset pricing", November 1986.
87/03	David GAUTSCHI and Vithala RAO	"A methodology for specification and aggregation in product concept testing", Revised Version: January 1987.
87/04	Sumantra GHOSHAL and Christopher BARTLETT	"Organizing for innovations: case of the multinational corporation", February 1987.
87/05	Arnoud DE MEYER and Kasra FERDOWS	"Managerial focal points in manufacturing strategy", February 1987.
87/06	Arun K. JAIN, Christian PINSON and Naresh K. MALHOTRA	"Customer loyalty as a construct in the marketing of banking services", July 1986.
87/07	Rolf BANZ and Gabriel HAWAWINI	"Equity pricing and stock market anomalies", February 1987.
87/08	Manfred KETS DE VRIES	"Leaders who can't manage", February 1987.
87/09	Lister VICKERY, Mark PILKINGTON and Paul READ	"Entrepreneurial activities of European MBAs", March 1987.
87/10	André LAURENT	"A cultural view of organizational change", March 1987
87/11	Robert FILDES and Spyros MAKRIDAKIS	"Forecasting and loss functions", March 1987.
87/12	Fernando BARTOLOME and André LAURENT	"The Janus Head: learning from the superior and subordinate faces of the manager's job", April 1987.
87/13	Sumantra GHOSHAL and Nitin NOHRIA	"Multinational corporations as differentiated networks", April 1987.
87/14	Landis GABEL	"Product Standards and Competitive Strategy: An Analysis of the Principles", May 1987.
87/15	Spyros MAKRIDAKIS	"METAFORECASTING: Ways of improving Forecasting. Accuracy and Usefulness", May 1987.
87/16	Susan SCHNEIDER and Roger DUNBAR	"Takeover attempts: what does the language tell us?, June 1987.
87/17	André LAURENT and Fernando BARTOLOME	"Managers' cognitive maps for upward and downward relationships", June 1987.
87/18	Reinhard ANGELMAR and Christoph LIEBSCHER	"Patents and the European biotechnology lag: a study of large European pharmaceutical firms", June 1987.

87/19	David BEGG and Charles WYPLOSZ	"Why the EMS? Dynamic games and the equilibrium Policy Regime, May 1987.

EURO-ASIA CENTRE

CAMPUS INSEAD

Boulevard de Constance
77309 Fontainebleau Cedex, France
Telephone (1) 60 72 40 40
Telecopy (1) 60 72 40 49
Telex 690389

EAC RESEARCH PAPERS

EAC RESEARCH PAPERS

(Academic papers based on the research of EAC Faculty and research staff)

1. LASSERRE Philippe (Research Paper n° 1)
A contribution to the study of entrepreneurship development in Indonesia. 1980.

2. BOISOT Max and LASSERRE Philippe (Research Paper n° 2)
The transfer of technology from European to ASEAN entreprises: strategies and practices in the chemical and pharmaceutical sectors. 1980.

3. AMAKO Tetsuo (Research Paper n° 3)
Possibilité d'un transfert à l'étranger des techniques japonaises de gestion du personnel: le cas français. 1982.

4. SCHUTTE Hellmut (Research Paper n° 8)
Wirtschaftliche Kooperation zwischen den ASEAN - Ländern und Nordrhein-Westfalen - Hemmungsfaktoren und Chancen für die deutsche Wirtschaft. 1983.

5. ISHIYAMA Yoshihide (Research Paper n° 14)
The political economy of liberalisation of the financial system in Japan. 1984.

6. LASSERRE Philippe (Research Paper n° 17)
Singapour comme centre régional. L'expérience d'entreprises françaises. 1985.

7. Von KIRCHBACH Friedrich (Research Paper n° 18)
Patterns of export channels to developing Asia. 1984.

8. MITTER Rajan (Research Paper n° 19)
A survey of European business in India. 1984.

9. CHAPON Marie-Claude (Research Paper n° 22)
Stratégies des entreprises japonaises en Afrique. 1985.