GENERALISING ABOUT UNIVARIATE FORECASTING METHODS: FURTHER EMPIRICAL EVIDENCE

by
R. FILDES*
M.HIBON**
S. MAKRIDAKIS†
and
N. MEADE††

96/70/TM

* Professor at Lancaster University, UK

** Research Associate at INSEAD, Boulevard de Constance, Fontainebleau 77305 Cedex, France.

† Research Professor of Decision Sciences at INSEAD, Boulevard de Constance, Fontainebleau 77305 Cedex, France.

†† Professor at Imperial College, London, England.

GENERALISING ABOUT UNIVARIATE FORECASTING METHODS: FURTHER EMPIRICAL EVIDENCE

Robert Fildes[1], Michèle Hibon[2], Spyros Makridakis[2] and Nigel Meade[3]

Revised Version, September 1996

Correspondence address:

Dr Nigel Meade

The Management School

Imperial College

London SW7 2PG

United Kingdom

Tel: (44) 0171 594 9116

Fax:(44) 0171 823 7685

e mail nme@ic.ac.uk

1 Lancaster University, UK
2 INSEAD, Fontainebleau, France
3 Imperial College, London, UK.

GENERALISING ABOUT UNIVARIATE FORECASTING METHODS: FURTHER EMPIRICAL EVIDENCE

Abstract

This paper extends the empirical evidence on the forecasting accuracy of extrapolative methods. The robustness of the major conclusions of the M-Competition data is examined in the context of the telecommunications data of Fildes (1992). The performance of Robust trend, found to be a successful method for forecasting the telecommunications data by Fildes, is compared with that of other successful methods using the M-Competition data. Although it is established that the structure of the telecommunications data is more homogeneous than that of the M-Competition data, the major conclusions of the M-Competition continue to hold for this new data set. In addition, while the Robust Trend method is confirmed to be the best performing method for the telecommunications data, for the 1001 M-Competition series, this method is outperformed by methods such as Single or Damped Smoothing. However, the performance of smoothing methods is shown to depend on how the smoothing parameters are estimated. Optimisation at each time origin is shown to be superior to optimisation at the first time origin, which in turn is shown to be superior to arbitrary (literature based) fixed values. In contrast to the last point, a data based choice of fixed smoothing constants from a cross-sectional study of the time series was found to perform well.

Keywords: Comparative methods - time series: Univariate / Time series - univariate: ARIMA / Estimation - robust / Time series - univariate: exponential smoothing / M-Competition.

GENERALISING ABOUT UNIVARIATE FORECASTING METHODS:

FURTHER EMPIRICAL EVIDENCE

The 1001 time series used in the M-Competition (Makridakis et al., 1982) have become a recognised collection of test data for the evaluation of forecasting methods. The four major conclusions of the M-Competition can be summarized as follows:

a. Statistically sophisticated or complex methods do not necessarily produce more accurate forecasts than simpler ones.
b. The rankings of the performance of the various methods varies according to the accuracy measure being used.
c. The accuracy of the combination of various methods outperforms, on average, the individual methods being combined and does well in comparison with other methods.
d. The performance of the various methods depends upon the length of the forecasting horizon.

In an examination of a set of 263 telecommunications related time series, Fildes (1992) drew these conclusions:

e. The characteristics of the data series are an important factor in determining relative performance between methods.
f. A method specifically derived to model and forecast the telecommunications series, Robust Trend (Grambsch and Stahel, 1990), performed best according to all the relevant error measures used, (outperforming several exponential smoothing methods which had performed well in the M-Competition).
g. Sampling variability of performance measures rendered comparisons based on single time series unreliable; comparisons based on multiple time origins, were recommended.

The studies of both sets of time series provide evidence of the accuracy of subsets of extrapolative forecasting methods. However the sets of time series are different and the subsets of forecasting methods considered differ in some important respects. The objective of this research is to consolidate and supplement the earlier evidence from the studies of both sets of time series.

The structure of the paper is as follows:

1. The two sets of time series are described and the background to the studies of each data set is given. In order to examine conclusion (e), summary statistics describing the time series are computed and used to describe the differences between the data sets. The further analyses needed to supplement the available published evidence are clearly identified.

2. The validity of the conclusions of the M-Competition (a, b, c, d) is examined using the telecommunications data. The value of these conclusions would be diminished if they were shown to be dependent on the set of time series used. With particular reference to (a), the range of models used with the telecommunications data is extended to include possibly appropriate candidates that have been applied to the M-Competition data. In the light of conclusion (g), multiple time origins were used for comparisons.

3. The performance of the Robust Trend model is evaluated using the M-Competition data. Here conclusion (f) is examined; since this model was derived specifically for the telecommunications data, it is of interest to discover whether its superiority is due to the special nature of these data or whether it will perform well on other time series.

4. The effect of changing the frequency of updating of smoothing parameters on the forecasting performance of exponential smoothing methods is examined. This is of interest since the performances of exponential smoothing methods are important benchmarks in most studies of forecasting accuracy.

The paper concludes with a discussion of the findings and suggestions for further avenues of research.

1. Data Sets[1], Error Measures and the Choice of Forecasting Methods

The M-Competition data set is 1001 time series collected from a variety of sources and covering micro, macro, and demographic data. These 1001 series were made up of 181 yearly, 203 quarterly and 617 monthly series. Makridakis et al., (1982) examined the post-sample forecasting accuracy of all major extrapolative (time series) methods, available at that time. The conclusions mentioned have been replicated by many researchers in three ways. First, the calculations on which the study was based were re-examined and their accuracy verified (Simmons, 1986). Second, new methods have been introduced and the results obtained have been found to agree with the conclusions of the M-Competition (Clemen, 1989; Koehler and Murphree, 1988; Geurts and Kelly, 1986). Tashman and Kruk (1996) examine the effect of selecting an appropriate exponential smoothing model for a series. They show that selection, rather than using all models for all series, may lead to a more accurate portrayal of differences in accuracy. Unfortunately, the various model selection criteria are prone to inconsistency. Finally, additional studies using new data series have concurred with these conclusions (Armstrong and Collopy, 1992; Makridakis et al., 1993). A full discussion of the impact of these empirical studies is given in Fildes and Makridakis (1995).

The telecommunications data set examined by Fildes (1992) is taken from one company and refers to one US state. Each series is monthly and the variable observed is the number of a particular type of circuit in service at a particular switching office (wire centre); hence the data is inherently non-negative. Forecasts of these variables are needed up to at least 12 months ahead for investment and installation planning. There are 263 time series in the set. According to Fildes, the series are characterized by no seasonality, low randomness, occasional outliers and a

[1] The M-Competition and the Telecommunications data sets are publicly available from the OR-Library (website: *http:/mscmga.ms.ic.ac.uk/info.html* or via anonymous ftp from *mscmga.ms.ic.ac.uk*).

downward sloping trend.

A systematic sample of four time series from each data set is shown in Figure 1. The downward trend is common to the four telecommunications series, with dramatic step changes apparent in series 104. Three out of the four M-Competition series shown exhibit far more variation about their trend than do the telecommunications series. In order to put this comparison on a firmer footing, some summary statistics have been computed to describe the time series from both data sets. All available values of the time series were used and the M-Competition series were deseasonalised (where necessary). Firstly, to remove the effects of series length and units of measurement, each series is standardised. The observed time series $x_1, \ldots, x_N$ is standardised to

$$y_t = \left(\frac{x_t - \min(x_1, \ldots, x_N)}{\max(x_1, \ldots, x_N) - \min(x_1, \ldots, x_N)} \right) . N$$

The difference $z_t = (y_t - y_{t-1})$ is computed. The effect of the standardisation is that for a deterministic linear trend, $E(Z_t) = \pm 1$ and $V(Z_t) = 0$. Since outliers distort average measures of trend and variation, outliers were identified and counted. (An impulse in x_t would generate two successive potential outliers in trend of opposite signs.) If the upper and lower quartiles of z_t are U_z and L_z respectively, an observation is an outlier if:

$$z_t < L_z - 1.5\,(U_z - L_z) \qquad \text{or, if} \qquad z_t > U_z + 1.5\,(U_z - L_z).$$

The mean and standard deviation of the non-outlying observations of z_t are computed. Histograms of these means and standard deviations are shown in Figure 2, along with a histogram of the proportion of observations classified as outliers. A plot of the most extreme series, (always from the M-Competition data set) is shown superimposed on each histogram. The contrasts between the two sets of time series are clear for each measure. The trends of telecommunications time series are more closely clustered and have negative trends (apart from three series). The average trends of the M-Competition series are more widely dispersed; their modal trend is 1.2.

Figure 1a. Telecomm.s Series 40 & 104

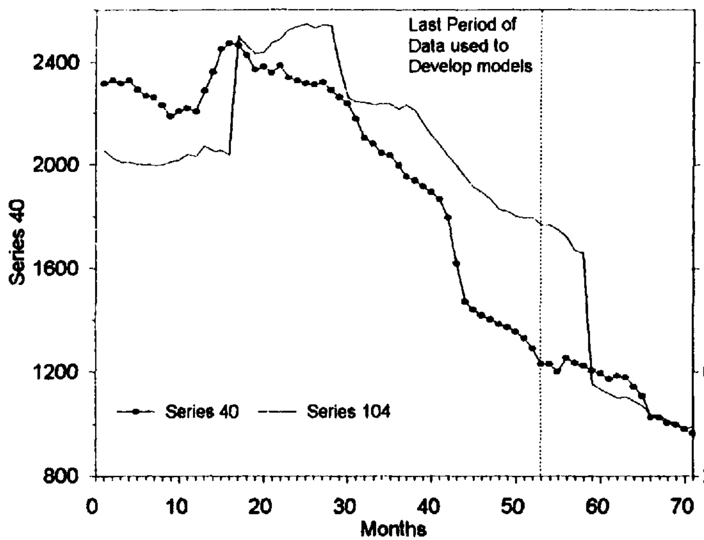


Figure 1c. M-Comp. Series 172 & 422

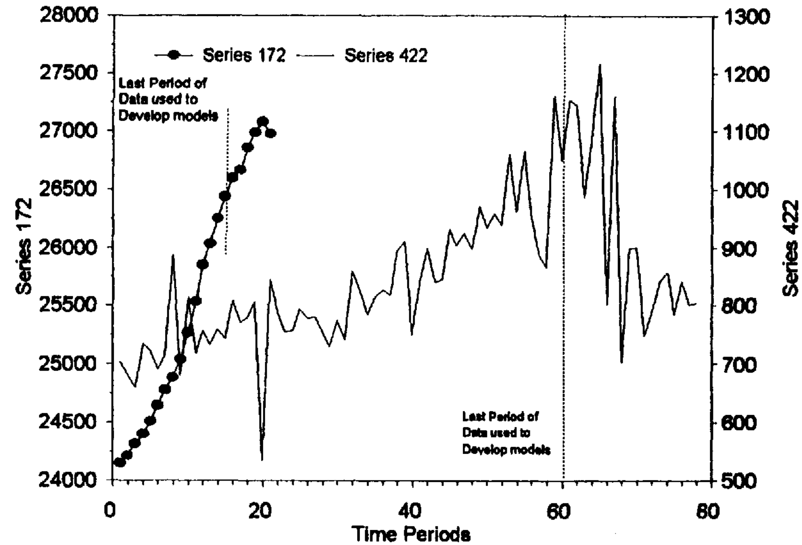


Figure 1b. Telecom.s Series 168 & 232

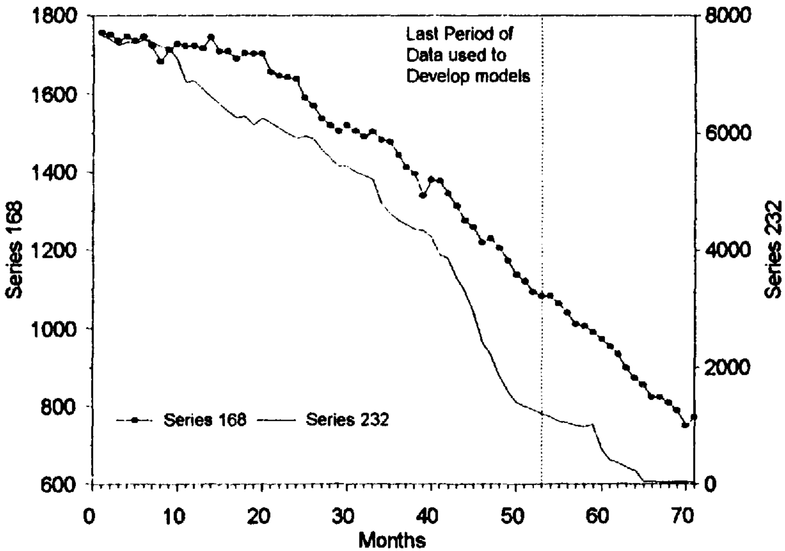


Figure 1d. M-Comp. Series 672 & 922

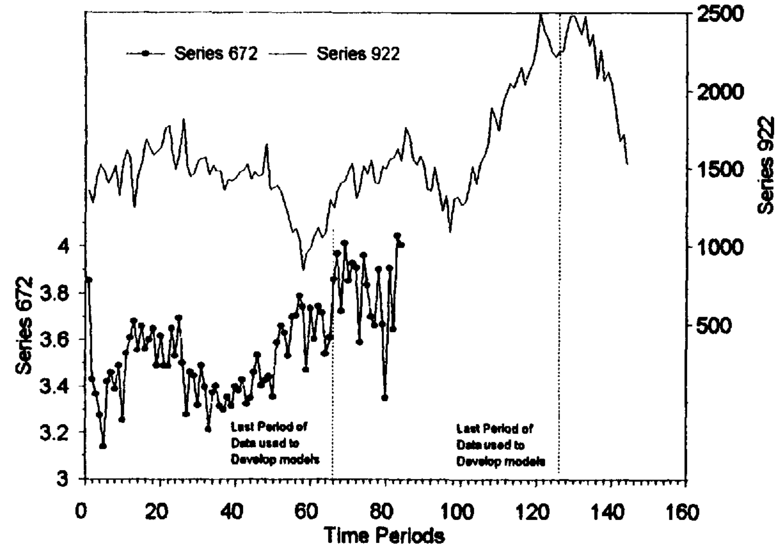
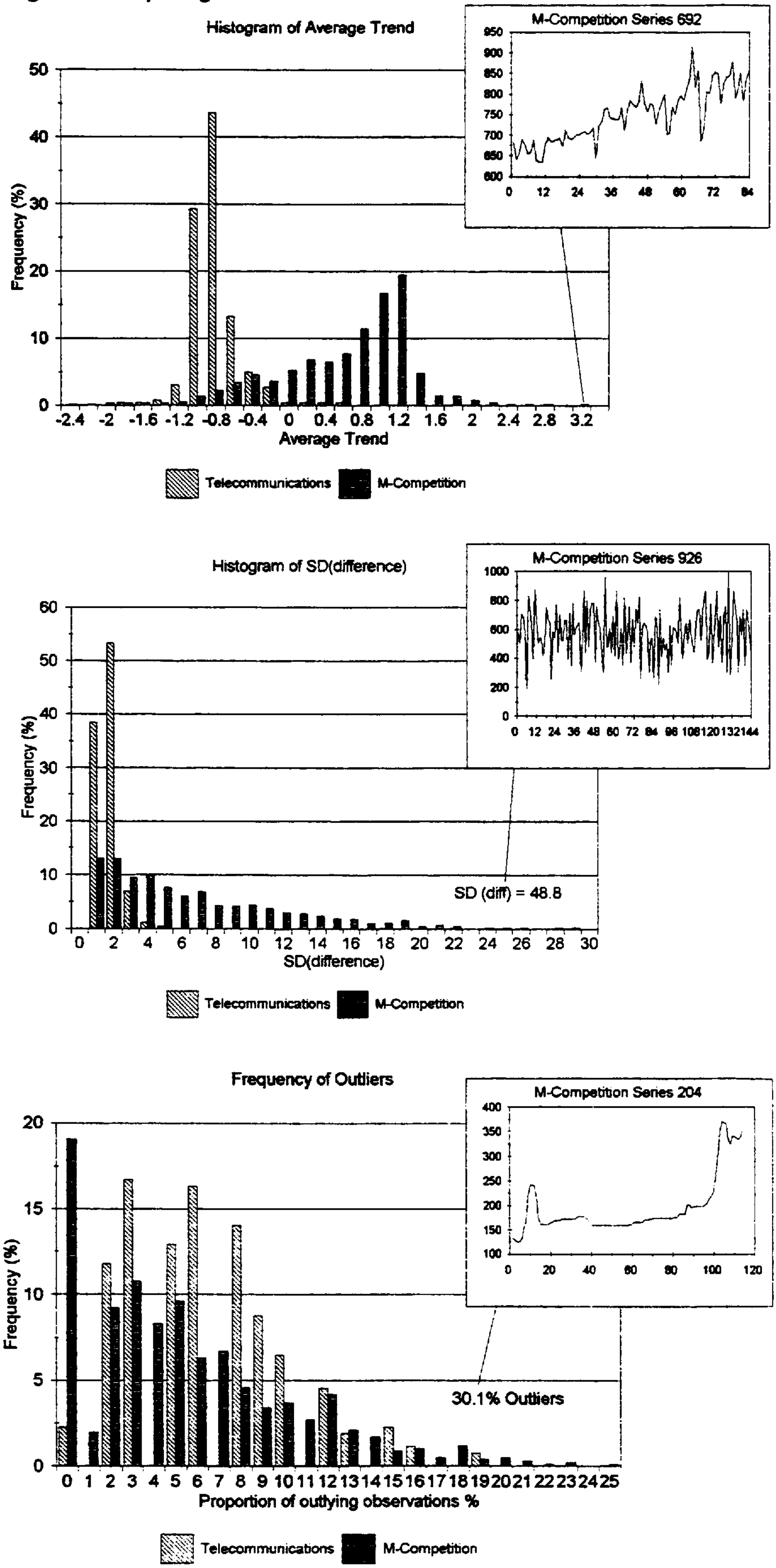


Figure 2. Histograms comparing the two sets of time series



Since, under this standardisation regime, a deterministic trend gives a value of unity, values greater than unity can only be achieved with at least one outlier. The standard deviations of the telecommunications series are tightly clustered with low values, confirming the comment of low randomness by Fildes. The M-Competition series exhibit a wide range of values; most of the M-Competition series have a far greater random component than the telecommunications data. Nearly all the telecommunications series exhibit at least one outlier, whereas 20% of M-Competition time series exhibit no outliers. (The example shown, M-Competition Series 204, possesses so many outliers because over half the values of z_t were zero.) In terms of the similarity of trend and level of random variation about trend, it is clear that the structure of the telecommunications data is more homogeneous than that of the M-Competition data.

Having examined the differences between the data sets and before looking at the effect of these differences on forecasting performance of different methods, it is appropriate to examine how performance is measured. The properties of different forecasting performance measures have been investigated by Armstrong and Collopy (1992). The error measures which will be used later are listed below with a summary of Armstrong and Collopy's comments (A&C). It is also noted whether the measure was used in the M-Competition.

i) MAPE: Mean absolute percentage error-used in the M-Competition: (A&C). A satisfactory comparative measure provided no large (percentage) errors are expected. This measure is obviously subject to distortion if the actual values are close to zero.

ii) MdAPE: Median absolute percentage error-used in the M-Competition: (A&C). Recommended for selection between forecasting methods when many series are available.

iii) % Better: The proportion of forecasts where one method is better than another - used in the M-Competition: (A&C). This measure is unaffected by outliers.

iv) Average Rankings: This is the average ranking of the absolute percentage error from each method for each horizon - used in the M-Competition

v) Geometric Mean RAE: Geometric relative absolute error: (A&C). This measure is recommended for calibrating the parameters of a particular method.

vi) Median RAE: Median relative absolute error: (A&C). Recommended for selection between forecasting methods when a small number of time series is available.

The use of mean square errors has been criticised (see for example, Chatfield, 1992) for its scale dependence and its sensitivity to outliers, so this measure was not included in this analysis.

In order to identify forecasting methods that may perform well on the telecommunications data, evidence can be sought from the analysis of M-Competition. The nearest sub-categories of the M-Competition data to the telecommunications data are those where seasonality does not play a part, that is yearly and non-seasonal data. Gardner and McKenzie (1985) noted that most of the yearly series exhibited a strong trend. The performance summary shown in Table 1 serves to identify methods that may be appropriate.

Table 1. Summary of performance extracted from the analysis of the M-Competition

Method	Performance Criterion									
	MdAPE				MAPE				Average Ranking	
	Yearly		Non-seasonal		Yearly		Non-seasonal		Yearly	Non-seasonal
	1-4	1-6	1-4	1-18	1-4	1-6	1-4	1-18	Overall	Overall
Naive1	10.2	12.9	7.5	10.9	13.6	17.1	14.7	21.1	15.06	12.85
Single Exp.	9.4	11.5	7.2	10.6	13.1	16.9	14.1	20.2	16.01	12.42
Holt	3.6	5.1	5.5	10.3	10.2	12.7	13.8	22.2	9.15	11.36
ARIMA	5.2	7.8	4.8	10.5	12.6	16.0	12.5	19.5	11.48	11.31
ARARMA	4.8	6.2	5.0	9.8	11.0	13.8	11.8	16.3	10.52	10.98
Combination A	6.4	7.7	6.0	10.0	10.8	13.5	12.1	20.3	11.32	10.90

Source: Makridakis et al, 1982

The ARARMA model performed well in the M-Competition, achieving the lowest MAPE for longer horizons for all 111 series and this is reflected in the subset of these data sets described above. The performance of the ARIMA method on non-seasonal data in the M-Competition was good (in terms of MdAPE) for the shorter horizons (up to 6 periods) where it out-performed ARARMA; its performance deteriorated with longer horizons where ARARMA dominated. This raises questions about the use of differencing to model trends in comparison with the autoregressive filter used in the ARARMA model. Holt's method does well in the MAPE and rankings measures for yearly data. Damped Exponential Smoothing or Holt's-D was proposed by Gardner and McKenzie (1985). This method was evaluated on the M-Competition data and performed better than Holt's method. The essential difference between the two methods is that Holt's method extrapolates a trend linearly and the damped version extrapolates a trend which decays towards a constant level.

In his analysis of the telecommunications data, Fildes (1992) used the Robust Trend (described in appendix 1), a Kalman filter method, Holt's method and Damped Smoothing. Robust Trend dominated the Kalman filter convincingly, so that method is not considered further here. Robust Trend also dominated the two smoothing methods.

The superior performance of the Damped Smoothing method on the M-Competition data gives additional insight to the structures of these time series. Gardner and McKenzie give the equivalent ARIMA forms of the exponential smoothing methods. The Damped Smoothing model is the most general, an ARIMA (1, 1, 2)

$$(1 - \phi B)(1 - B) X_t = (1 - \theta_1 B - \theta_2 B^2) a_t$$

The other models are special cases of Damped, Holt's model is ARIMA (0,2,2) achieved by setting ϕ=1, Single Smoothing is ARIMA (0, 1, 1) obtained by setting $\phi = 0$ and $\theta_2 = 0$. With these models, the trend is stochastic; previous errors made by the model adjust the estimated trend. The high values of the standard deviation of the period to period changes in the M-Competition data seen earlier are consistent with the random component implied:

$$\frac{(1 - \theta_1 B - \theta_2 B^2)}{(1 - \phi B)} a_t ;$$

which may have a variance substantially larger than $V(a_t)$. The telecommunications data have been seen to exhibit a combination of negative trend, low variation about the trend and several outliers. The Robust Trend model, built for these data, has two facets; its robustness which is gained by using estimation procedures unaffected by outliers; and its model for trend. In ARIMA notation, the underlying model is

$$(1 - B) X_t = \Theta_0 + a_t ,$$

an ARIMA (0, 1, 0) with a constant, i.e. a random walk with a deterministic trend. The comparatively low level of variation about the trend is thus assumed to be white noise. The robustness of the estimate of the deterministic trend parameter, Θ_0 is given by using the median of the differenced data (subject to the adjustments shown in appendix 1) rather than the arithmetic mean. This protects the estimate of the trend parameter from being contaminated by the outliers that are particularly common in the telecommunications data.

A summary of the forecasting methods identified as being of possible interest is shown in Table 2. Their previous applications to the two data sets are identified and the analyses likely to give further evidence about forecasting performance are shown. In further analysis of the telecommunications data, out of the methods listed in Table 2, we can discount the two naive

Table 2. Forecasting Methods used

Method	Reference	Used previously with the M-Competition data	Used previously with the Telecom.s data	Referred to in section number:
Naive/Simple Methods				
Naive 1 (Random Walk) or Naive 2 for seasonal data	Makridakis et al., (1982)	Yes	No	3
Single Exponential Smoothing	Makridakis et al., (1982)	Yes	No	3, 4
Explicit Trend Models				
Holt's Exponential Smoothing	Makridakis et al., (1982)	Yes	Yes [2]	2, 3, 4
Damped Exponential Smoothing	Gardner and McKenzie, (1985)	Yes	Yes [2]	2, 3, 4
Robust Trend	Grambsch and Stahel, (1990)	No [1]	Yes [2]	2, 3
***Sophisticated* Methods**				
ARIMA (represented by an automatic approach)	Stellwagen and Goodrich, (1991)	Previous ARIMA analyses have used 111 series [1]	No [1]	2, 3
ARARMA	Parzen, (1982)	Yes	No [1]	2

[1] Additional analyses carried out in this study

[2] A reworking of analysis in Fildes (1992) due to reasons specified in the text

NOTE the ordering of methods used here, (increasing complexity) will be used in succeeding tables methods as they will not model a trend, the two remaining models are ARARMA and ARIMA. Self evidently, since an appropriate ARIMA model has been used to describe the telecommunications data, an ARIMA model is eligible. Since, notationally at least, ARARMA is a generalisation of ARIMA, it also must be considered an eligible model. In addition, ARARMA is designed to model long memory processes and it is conceivable that the trend in the data may be so modelled. Both ARIMA and ARARMA are theoretically capable of modelling the telecommunications data; the possible problems may arise in estimation. The high proportion of outliers in the time series of this data set may undermine the identification and estimation process. Wu et al (1993) and Ledolter (1989) found that the presence of outliers, unless very close to the forecast origin, had little effect on forecast accuracy. The main effect was to cause excessively wide prediction intervals. The performance of these methods on the telecommunications data set is examined in section 2. In further analysis of the M-Competition data, the performance of Robust Trend will be evaluated in section 3. Its overall performance will be determined by the balance of benefits brought by the method. Its robust estimation process offers benefits since the above analysis of this data shows the existence of at least one outlier in about 80% of the series. Its representation of all variation about a trend as white noise limits its capability of modelling the random components.

2. Validation of the M-Competition conclusions using the telecommunications data

This section examines whether the four major conclusions (a, b, c, d) of the M-Competition hold for the telecommunications data. It is a strong test of their generalisability, in that the two data sets differ substantially as we have just shown. Two of the 263 telecommunications series were dropped from this analysis because both had many zero values. Following Fildes (1992), this evaluation uses five time origins, in contrast with the single time origin used in the M-Competition. Instead of comparing the post-sample forecasting accuracies at the end of the data, assuming that a single set of forecasts is being made, five sets of rolling forecasts are compared. Each time series has 71 monthly observations and the five sets of forecast evaluations are made with time origins in months: 23, 31, 38, 45 and 53. As usual, when we make post-sample comparisons we assume that only information up to the time origin is available. Consequently the forecasts made are compared with the actual values - not used to develop the model - in order to determine post-sample accuracies and evaluate the performance of the various methods. If the downward sloping trend leads to negative forecasts, theses forecasts are reset to zero.

The robustness of conclusions a, b, c and d will be examined using the telecommunications data in the following three subsections.

i) Sophisticated vs. Simple Methods (Conclusion a)

In the analysis to be described, the smoothing parameters for all exponential smoothing methods were optimised for each time origin (Fildes, 1992, optimised once at time origin 23 and applied these values when forecasting for all subsequent time origins). The issue of the frequency of optimisation is discussed later in section 4. Figure 3 shows the Mean Absolute Percentage Error (MAPE) for five forecasting methods when the evaluation was done in exactly the same way as in the M-Competition (that is, observations 1 to 53 were used to develop a model and then 1 to 18 step ahead forecasts were made from a time origin of observation 53). Consequently the accuracy of these forecasts was evaluated with the 18 actual values available, but not used in developing the model. Figure 3 indicates that the Robust Trend outperforms all other methods

Figure 3. Comparison of forecasting performances for the Telecommunications data-Mean Absolute Percentage Error (MAPE) for time origin period 53.

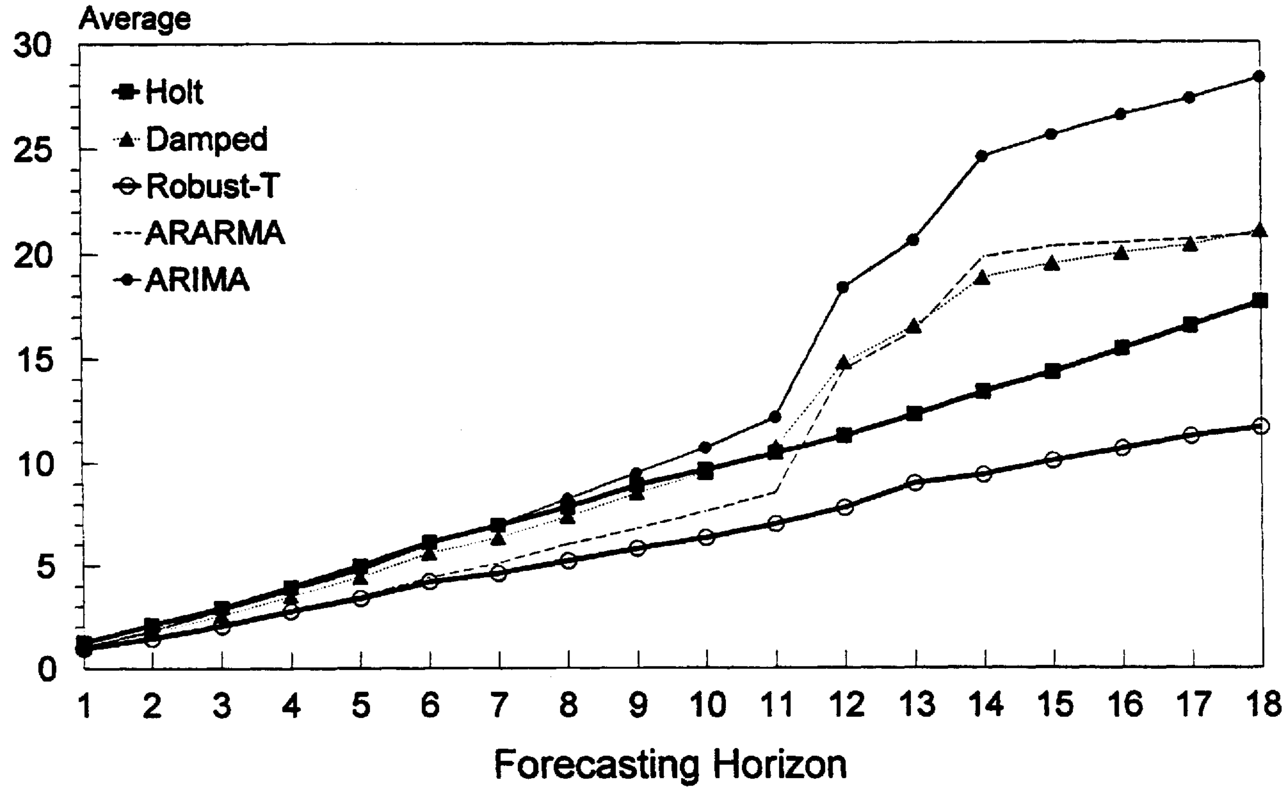
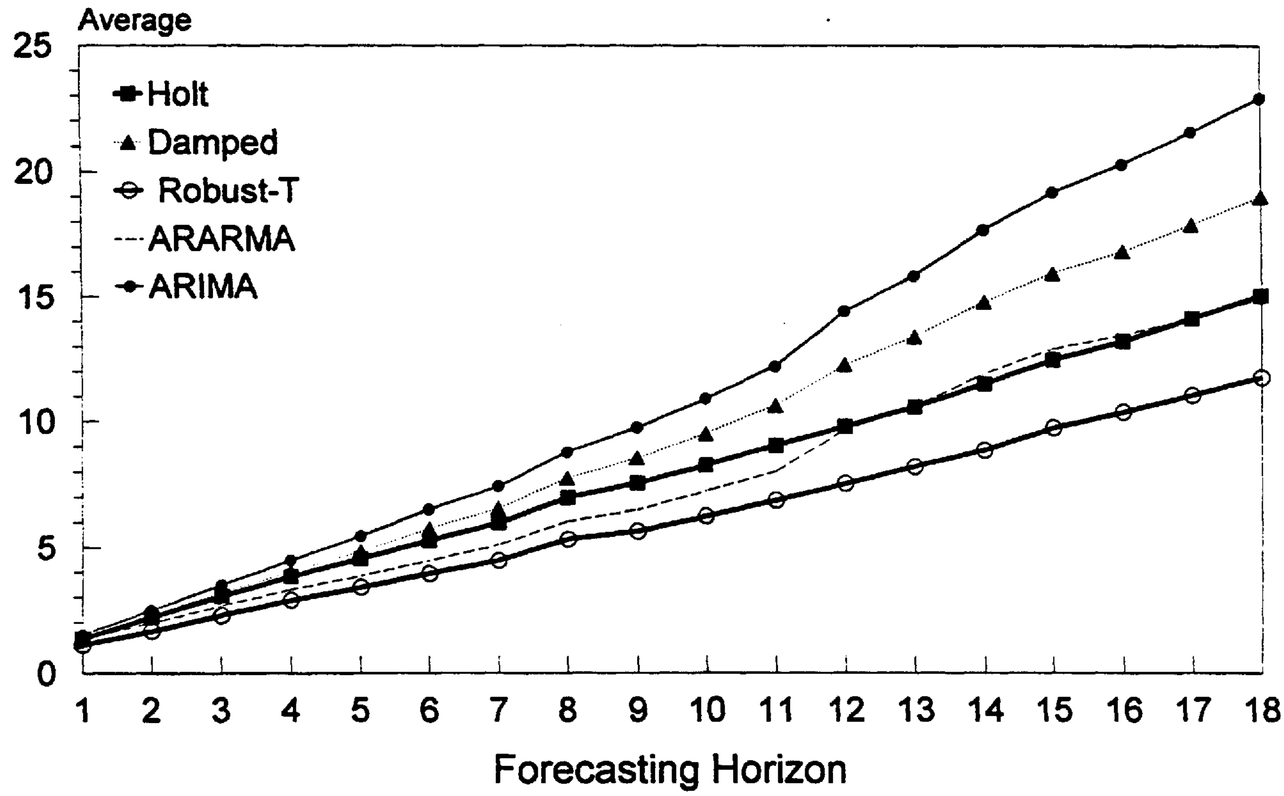


Figure 4. Comparison of forecasting performances for the Telecommunications data-Mean Absolute Percentage Error (MAPE) averaged over all time origins



in all forecasting horizons and consequently for the overall average[2] (see also Table 3). Holt's Smoothing is second best overall because of its superior performance at longer forecasting horizons (twelve to eighteen), while Parzen's ARARMA model (the most accurate method of the M-Competition in terms of MAPE) is third best. ARIMA models (the other sophisticated method with ARARMA) did worse than Holt's Smoothing and, in most cases not as well as Damped Smoothing. As discussed earlier, the methods of Naive and Single Smoothing are not appropriate for this data set because by the first time origin, month 23, the negative trend would be apparent to the forecaster. If such prior information was ignored and the strong performance of these two methods in the M- and M2-Competitions was taken into account, a forecaster choosing to adopt Single Smoothing instead of Robust Trend would have incurred losses in accuracy ranging from 18.9% at forecasting horizon 1 to 127.5% at horizon 18.

ii) Accuracy Measure Used and Length of Forecasting Horizon (Conclusions b & d)

Further details of the MAPEs for different time origins are given in Table 3. The average MAPE across time origins is summarised at the foot of the table and shown in more detail in Figure 4. As Fildes (1992) noted there is variation between the comparisons based at different time origins. Robust Trend dominates in all but one time origin. If one had used month 38 as a basis for analysis, one would have concluded that Holt's method dominated Robust Trend. For three time origins ARARMA is second best to Robust Trend, Holt's method is best or second best for the others (including time origin month 53 discussed above).

Unlike the M- and M2-Competitions Holt's Smoothing outperformed Damped for practically all ending periods/horizons because of the persistent negative trend that characterized most series. Let us change the focus from MAPE to the other measures of forecasting performance discussed.

---

2 There is a slight difference between the calculations of Robust Trend reported in Fildes (1992) and those of this study due to a minor programming error in the previous analysis.

Figure 5 shows the Median APE across all five time origins. Over longer horizons Holt's method performs better than ARARMA. This suggests that Holt's performance was characterised by outliers of high MAPEs for some series that influence the mean but not the median. In Figure 6, Robust Trend is taken as a benchmark and the percentage of occasions when it is better than the other methods is shown. The average rankings are shown in Figure 7 (the lower the rank the better the method). Holt's method and ARARMA perform similarly for most forecasting horizons. The relative absolute measures are considered next. They are shown in Table 4. The relative absolute error uses a 'no change' forecast as a basis of comparison, i.e. for a forecast with horizon, h, made at time t, it is:

$$\text{Relative Absolute Error(t,h)} = \frac{| \hat{X}_{t+h} - X_{t+h} |}{| X_t - X_{t+h} |}$$

Since the denominator can be zero, Armstrong and Collopy recommend Winsorizing the values, the minimum value used is 0.01, the maximum value is 10.0. In this data set, for more than half of the time series, X_{54} is equal to X_{53}. This leads to a median of 10 for the one step ahead forecasts for many methods for time origin 53. Since the values tabulated are an arithmetic mean over the five time origins, the values for one step ahead forecasts for both the median and geometric mean relative absolute error should be ignored. This quirk of the data (perhaps representing some unknown feature of the data collection process) highlights a weakness in the use of the relative absolute error measures, since the ranking of methods may be determined by the arbitrary choice of the Winsorizing values. (It also highlights the need to analyse more than one time origin.)

Table 5 summarises the differences in forecasting performance between the methods according to the criterion used. Although Robust Trend outperforms the other methods for practically all

Figure 5. Comparison of forecasting performances for the Telecommunications data - Median Absolute Percentage Error (MdAPE) over all five time origins.

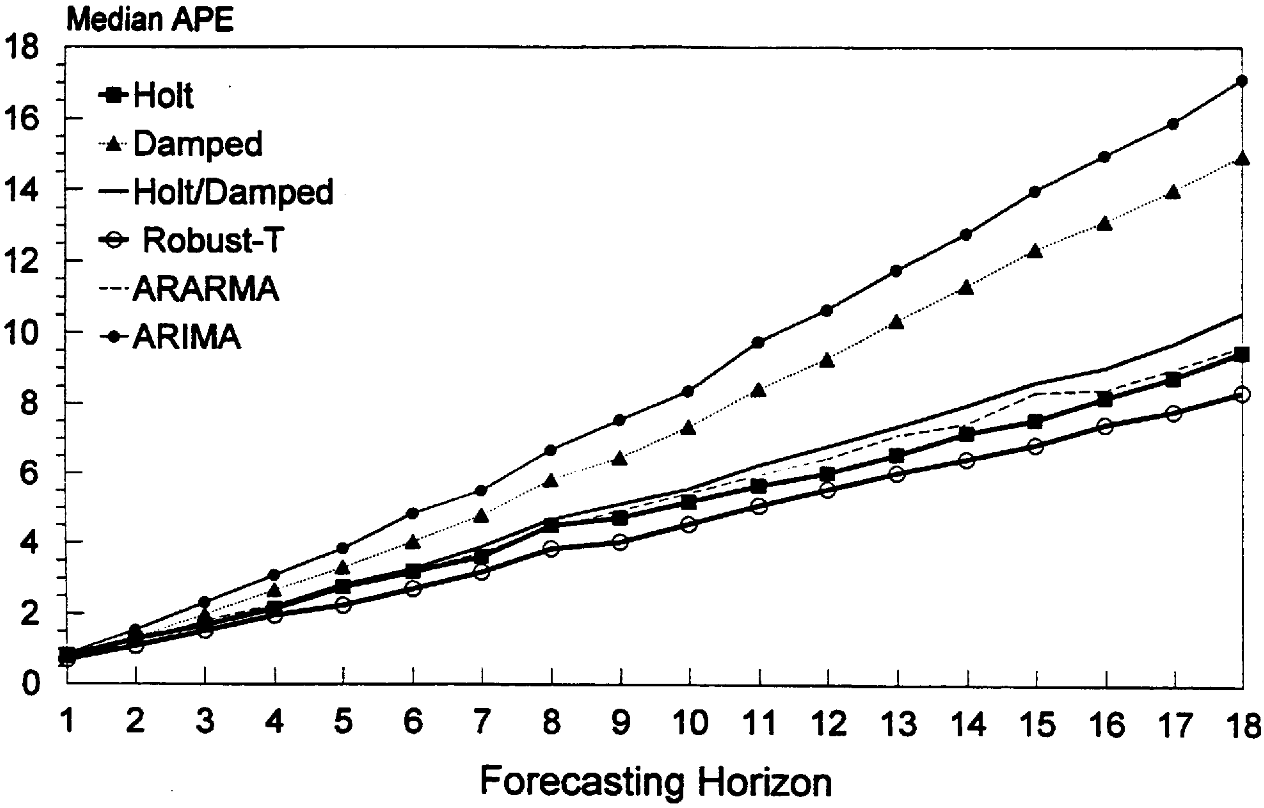


Figure 6. Comparison of forecasting performances for the Telecommunications data - % of times Robust Trend is better than method listed for all time origins

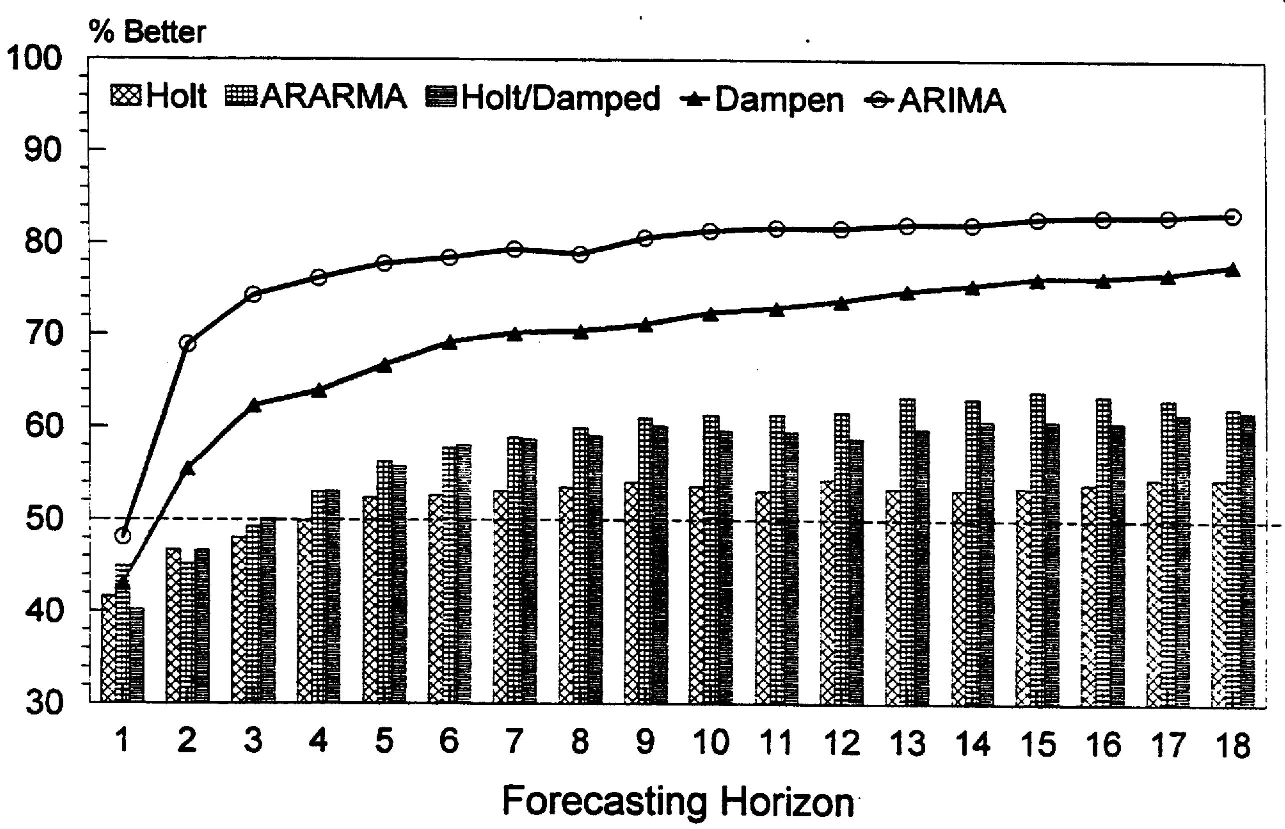


Figure 7. Comparison of forecasting performances for the Telecommunications data - Average Rankings over all five time origins

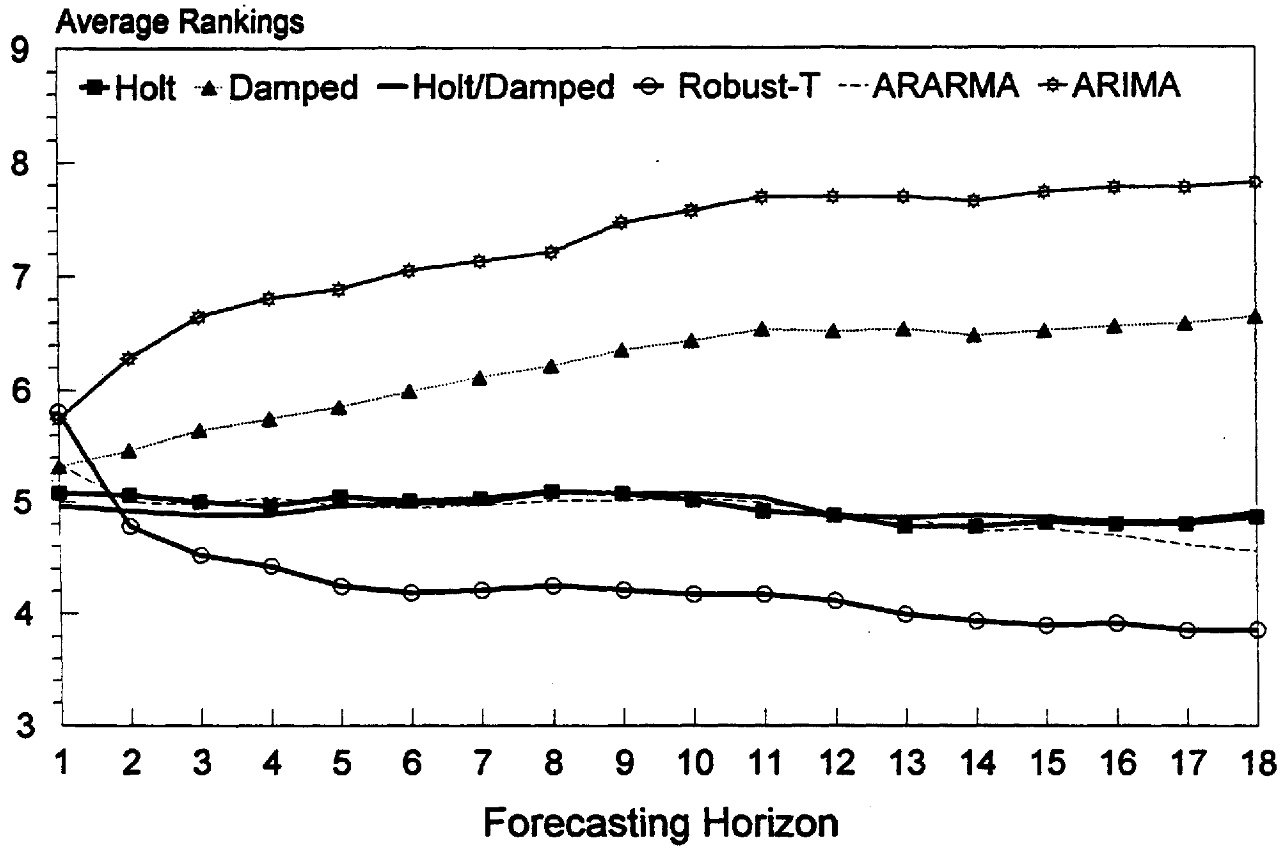


Table 3 The MAPEs of the different methods for different time origins

Time Origin	Method	Forecasting Horizon						
		1	6	12	18	1-6	1-12	1-18
23	Holt	1.46	6.79	11.30	16.20	4.43	6.99	9.37
	Damped	*1.25*	6.46	11.05	17.14	4.27	6.67	9.25
	Robust-T	***0.68***	***3.26***	***6.09***	***10.27***	***2.06***	***3.48***	***5.14***
	ARIMA	1.77	6.90	12.19	19.17	4.52	7.16	10.12
	ARARMA	1.60	*4.20*	*7.25*	*10.88*	*2.98*	*4.50*	*6.12*
31	Holt	*1.79*	5.30	10.77	16.66	3.59	6.09	8.80
	Damped	1.97	6.36	13.23	20.53	4.15	7.25	10.65
	Robust-T	***1.65***	***4.19***	***8.40***	***12.80***	***2.97***	***4.83***	***6.88***
	ARIMA	2.03	6.76	14.51	22.65	4.40	7.83	11.62
	ARARMA	1.95	*4.88*	*9.45*	*14.07*	*3.55*	*5.58*	*7.76*
38	Holt	***1.19***	***4.10***	***7.78***	***11.91***	***2.67***	***4.40***	***6.31***
	Damped	1.34	5.74	11.84	18.83	3.45	6.25	9.51
	Robust-T	*1.25*	*4.32*	*8.15*	*12.11*	*2.74*	*4.56*	*6.56*
	ARIMA	1.53	6.93	13.94	22.27	4.17	7.47	11.28
	ARARMA	1.28	4.64	8.90	13.84	2.97	4.94	7.24
45	Holt	1.10	*4.10*	*8.09*	*12.92*	*2.61*	*4.50*	*6.63*
	Damped	1.11	4.72	10.65	17.57	2.86	5.55	8.53
	Robust-T	***1.02***	***3.80***	***7.35***	***12.26***	***2.45***	***4.25***	***6.14***
	ARIMA	1.19	5.97	13.10	22.23	3.47	6.84	10.65
	ARARMA	*1.04*	4.31	8.64	15.03	2.65	4.84	7.26
53	Holt	1.25	6.09	*11.16*	*17.54*	3.54	6.32	*9.15*
	Damped	1.06	5.58	14.69	20.95	3.15	6.31	10.64
	Robust-T	***0.94***	***4.18***	***7.71***	***11.55***	***2.46***	***4.27***	***6.25***
	ARIMA	1.03	6.00	18.30	28.26	3.36	7.14	13.24
	ARARMA	*0.97*	*4.38*	14.37	20.82	*2.56*	*5.30*	10.08
Average	Holt	1.36	5.28	9.82	15.05	3.37	5.66	8.05
	Damped	*1.35*	5.77	12.29	19.00	3.58	6.41	9.71
	Robust-T	***1.11***	***3.95***	***7.54***	***11.80***	***2.54***	***4.28***	***6.19***
	ARIMA	1.51	6.51	14.41	22.92	3.99	7.29	11.38
	ARARMA	1.37	*4.48*	*9.72*	*14.93*	*2.94*	*5.03*	*7.69*

Lowest MAPE shown in ***bold italics***, second best shown in *italics*.

Table 4. The Relative Absolute Error measures

	Geometric Mean RAE			
Horizon	1*	6	12	18
Holt	*1.35*	*0.72*	***0.60***	*0.54*
Damped	1.39	0.81	0.78	0.77
Holt/Damped	1.36	0.75	0.66	0.60
Robust Trend	*1.35*	***0.71***	*0.61*	***0.52***
ARIMA	**1.33**	0.81	0.82	0.84
ARARMA	1.38	0.75	0.67	0.60
	Median RAE			
Horizon	1*	6	12	18
Holt	***1.16***	*0.76*	***0.64***	***0.59***
Damped	1.18	0.90	0.90	0.92
Holt/Damped	*1.17*	0.79	0.72	0.68
Robust Trend	*1.17*	***0.75***	0.66	*0.60*
ARIMA	2.80	0.97	0.98	0.99
ARARMA	1.77	0.78	*0.71*	0.65

* These values are distorted by the Winsorizing process - see text.

Table 5 Summary Rankings of Forecasting Performances

Criterion	Average MAPE	Median APE	% of times better than Robust Trend	Geometric Mean RAE		Median RAE
Horizon Method	18 months or cum. 1- 18 months	9 - 18 months	5 - 18 months	18 months	12 months	18 months
Holt	3	2	2	2	1	1
Damped	4	4	4	4	4	4
Robust Trend	1	1	1	1	2	2
ARIMA	5	5	5	5	5	5
ARARMA	2	3	3	3	3	3

accuracy measures used, the extent and consistency of such out-performance depends upon the specific measure being used. Furthermore, differences in performance between methods become more pronounced depending on the accuracy measures used. These orderings of forecasting methods should not be taken in isolation because they give no information about differentials in performance. For example, in Figure 5, the major difference is between the cluster of methods giving a MdAPE of around 8% and the cluster giving a MdAPE of around 16%. This consistently excellent performance of Robust Trend is rare in empirical comparisons and requires considering Robust Trend as another time series method that can contribute to improving forecasting accuracy when there is at least a consistent trend in the data.

Conclusion (d) of the M-Competition was that some methods are better for shorter forecasting horizons while others are more appropriate for longer ones. Evidence supporting this conclusion is shown in Table 3, for example, Damped Smoothing offers the second best MAPE for a one month horizon, but ARARMA is second best for longer horizons. Similar contrasts have been seen in Figures 3, 6 and 7 which show changing orderings as the horizon changes. Robust Trend does extremely well for longer forecasting horizons for MAPE and MdAPE while Holt improves its performance considerably with the remaining measures for long forecasting horizons. Damped trend does not do well for long forecasting horizons because the implicit assumption of a decaying trend is unfounded for the majority of these series. Robust Trend does not do well, according to the rankings criterion, for the first forecasting horizon while subsequently it does extremely well for longer horizons.

iii) The Combination of Holt and Damped Exponential Smoothing (Conclusion c)

Combination method A (used in the M-Competition), an unweighted average of six methods, is shown in Table 1 as a reminder of the benefits of combining forecasts from different methods (conclusion c). In this study the simple arithmetic average of Holt's and Damped Smoothing was calculated because of the simplicity and ease of running these two methods and combining their forecasts. The performance of the combination using most criteria has already appeared. The MdAPE is shown in Figure 5; the percentage of times Robust Trend is better than the combination is shown in Figure 6; its average ranking is shown in Figure 7 and its relative absolute error measures are given in Table 4.

Table 6 MAPEs for combined forecast and other methods

	Forecasting horizon			
Method	1	6	12	18
Holt	1.36	5.28	9.82	15.05
Damped	1.35	5.77	12.29	19.00
Holt/Damped	*1.31*	*5.24*	9.80	*14.74*
Robust-T	***1.11***	***3.95***	***7.54***	***11.80***
ARIMA	1.51	6.51	14.41	22.92
ARARMA	1.37	4.48	*9.72*	14.93

Table 6 shows the overall MAPE (average of all five ending periods) of Holt, Damped, their combination and the other methods (this partially duplicates the foot of Table 3). The combination outperforms the two methods being averaged across forecasting horizons. For all eighteen forecasting horizons the overall average of combining is 7.92% versus 8.05% for Holt (an improvement of 1.6%) and 9.71% for Damped (an improvement of 18.4%). This performance of combining Holt and Damped is puzzling since Damped is clearly a sub-optimal method as there is a consistent negative linear trend in the greater part of this set of 261 series. This point is revisited in section 3. Using MAPE as a criterion, combining also works better in relation to other methods being outperformed only by Robust Trend and having similar accuracy to ARARMA models. This performance of combining is consistent with that observed in the forecasting literature as well as in other areas (Clemen, 1989).

For median APE, Holt's method out-performs the combination for longer forecasting horizons; for average rankings there is little difference between Holt's and the combined method. The effect of the combining is likely to remove outlying cases of poor performance by either component, since (unless the trend parameter is zero or the damping parameter is one) the methods will produce different forecasts for longer horizons. This property would benefit the MAPE for the combined method, but not affect the median measure. Both relative absolute error measures show Holt's method out-performing the combined method for longer forecasting horizons. Taken together, these findings are not as strong as those observed in the M-Competition (Makridakis and Winkler, 1983; and Makridakis et al., 1984).

3. Performance of Robust Trend on M-Competition data

The Robust Trend and automatic ARIMA were applied to the M-Competition data. Results for the other methods tabulated are taken from Makridakis et al (1982), apart from Damped Exponential Smoothing which are taken from Gardner and McKenzie (1985). These results, in terms of mean and median APEs are shown in Table 7. An initial comment is that the values for MAPE for the M-Competition series are about 8 times larger for one step ahead forecasts than for the telecommunications data, the MdAPEs are about 5 times larger. For the 18 step ahead forecast, the MAPEs are about twice as large, for the MdAPEs are about 20% larger. This is a demonstration of the comparatively larger random component in the M-Competition data than that in the telecommunications data, as well as the effect of trend on accuracy for longer horizons.

The dominant method in this comparison is Damped Exponential Smoothing, which achieves the lowest error measure in all but one horizon shown in Table 7. The Robust Trend method performs similarly to Naive2, Single Exponential and Holt's Smoothing in terms of MAPE and

MdAPE. Using MAPE as a criterion, Robust Trend marginally outperforms ARIMA, except for short horizons. The position is exactly reversed using MdAPE.

The results for the 111 series in the M-Competition (those analysed by the sophisticated techniques) are shown in Table 8. The picture is not straightforward; the comments about the choice of method being determined by the forecasting horizon of interest are re-enforced here.

Table 7. MAPEs and MdAPEs for 1001 M-Competition time series

	Method	Forecasting Horizons			
		1	6	12	18
MAPE	Naive2	9.1	19.9	17.1	26.3
	Single	*8.6*	*19.6*	*16.9*	26.1
	Holt	8.7	21.6	23.9	48.3
	Damped	***8.3***	***17.9***	***16.7***	***21.7***
	Bayesian	11.2	21.0	18.9	28.3
	Robust-T	8.8	19.7	19.0	*24.3*
	ARIMA	*8.6*	20.8	20.4	40.3
MdAPE	Naive2	4.8	10.8	10.4	12.6
	Single	4.7	10.5	10.3	*12.5*
	Holt	4.5	***9.2***	9.9	13.6
	Damped	***4.2***	*9.3*	***9.3***	***11.9***
	Bayesian	5.1	10.0	10.1	13.1
	Robust-T	*4.3*	9.6	10.5	13.7
	ARIMA	4.5	9.7	*9.8*	13.0

For one period ahead forecasts, damped smoothing has the lowest MAPE and MdAPE and Robust trend is outperformed by all three exponential smoothing methods. For 18 period ahead forecasts, ARARMA dominates and the performance of Robust Trend is broadly similar to that of the exponential smoothing methods. An interesting by-product is the comparison between the automatic ARIMA approach used here with the original manual ARIMA models used in the M-Competition. The automatic approach produces better one period ahead APE measures, for longer horizons the MAPEs are worse, the MdAPEs are roughly similar. This may indicate that the manual forecasters were better at avoiding a few very poor forecasts which affect the MAPE but not the MdAPE. (See also Hill and Fildes, 1984 and Libert, 1984)

Examination of a series from each data set helps gain some insight in to the causes of the contrasting performance. In Figure 8, forecasts of series 40 (shown in full in Figure 1a) are shown for the last time origin. The Damped Smoothing forecast removes the trend almost immediately, Holt's Smoothing estimates the trend to be steeper than the Robust Trend estimate. In the short term, the Damped forecast is better, for longer horizons the Robust Trend is better.

Table 8. MAPEs and MdAPEs for 111 M-Competition time series

	Method	Forecasting Horizons			
		1	6	12	18
MAPE	Naive2	8.5	17.4	14.5	30.8
	Single	*7.8*	17.2	***13.6***	30.1
	Holt	7.9	17.8	16.4	34.4
	Damped	***7.6***	*15.9*	***13.6***	29.5
	Bayesian	10.3	17.1	16.1	30.6
	Robust Trend	8.4	16.9	15.3	*28.0*
	ARIMA (Auto)	8.6	19.3	17.3	39.0
	ARIMA (Man)	10.3	17.1	16.4	34.2
	ARARMA	10.6	***14.7***	13.7	***26.5***
MdAPE	Naive2	4.0	9.9	*7.3*	15.6
	Single	3.3	9.8	8.6	15.8
	Holt	3.4	9.0	7.4	15.0
	Damped	***2.8***	9.5	7.9	15.5
	Bayesian	5.0	9.3	8.8	*14.0*
	Robust Trend	3.5	***8.6***	9.4	15.1
	ARIMA (Auto)	*3.1*	9.3	7.7	16.9
	ARIMA (Man)	5.3	*8.8*	8.6	16.4
	ARARMA	4.8	9.0	***6.6***	***11.6***

Figure 8. Telecoms. Series 40
forecasts for time origin month 53

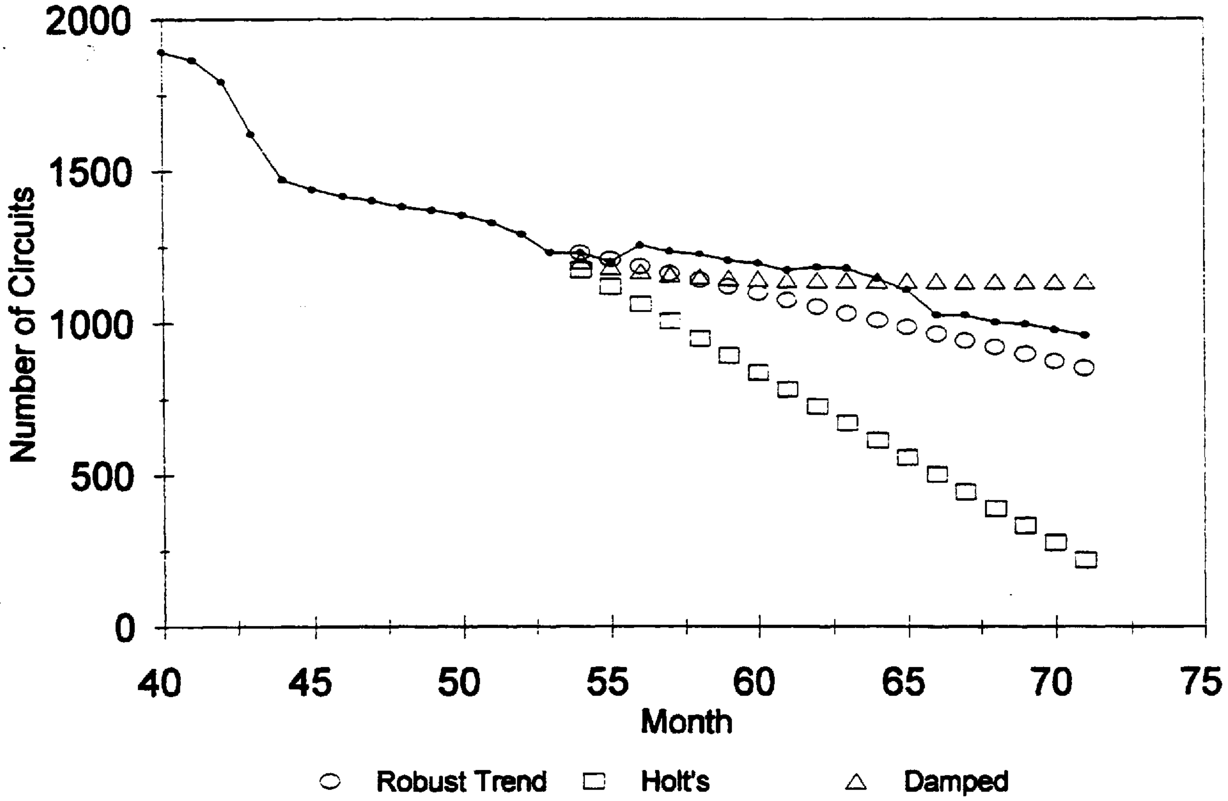
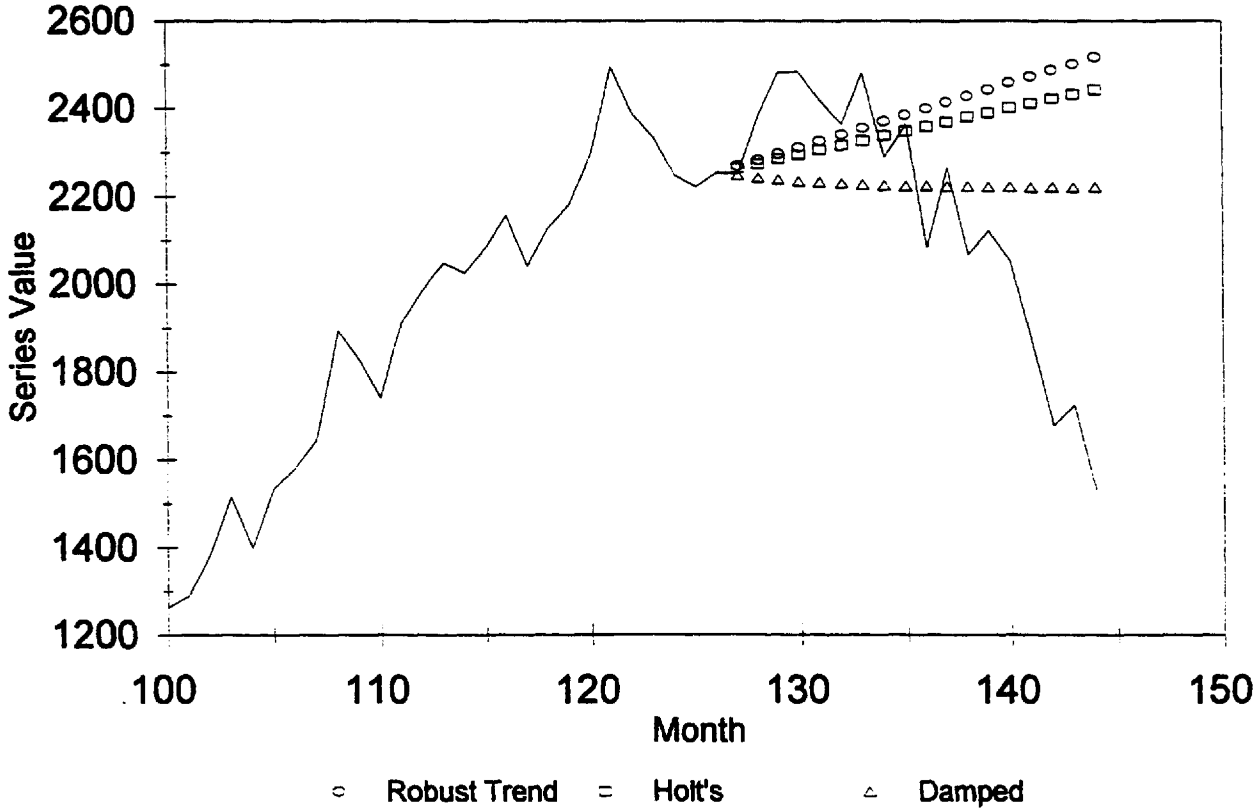


Figure 9. M-Competition Series 922
forecasts for time origin month 126



The combined forecast computed as a simple average of Damped and Holt's smoothing would be similar to the Robust Trend forecast. In Figure 9, the last 18 periods of one of the 111 M-Competition series (number 922 shown in full in Figure1d) are forecast using the same three methods. For this origin, the trend is estimated to be slightly negative by Damped smoothing. Holt's smoothing and Robust Trend both extrapolate the trend upwards, the latter more steeply, forecasting well for short horizons; then suffering disastrously as the data fall sharply. This property of Damped smoothing to provide an 'each way bet' by forecasting no trend is an advantage over a method like Robust Trend. The underlying difference between the two series is that in the telecommunications data, the trend observed up to the time origin is likely to persist; in the M-Competition data it is not. The probability of a trend persisting determines the likely ranking between the performances of Robust Trend and Damped Smoothing.

4. Smoothing methods: Three ways of dealing with their parameters

The choice of values for smoothing parameters in exponential smoothing can be made in a number of distinct ways. In earlier comparisons, the question of whether the particular variant used affects performance has not been examined. In this section, the performance of three exponential smoothing methods, Single, Holt and Damped is examined under **three** operating conditions. These conditions vary in the amount of information drawn from the time series used to set the model parameters.

- **Arbitrary values**. In this case no information from the series is used, the parameter values were set arbitrarily to certain values suggested in the forecasting literature (many programs, for example SAS, expect the value of smoothing parameters to be a user input). For Single Exponential Smoothing, Brown (1963) suggests that the parameter α is set to 0.1. Coutie et al (1964) recommend values of 0.1 and 0.01 for the smoothing parameters of level and trend respectively in Holt's method. This recommendation is extended to Damped Exponential Smoothing with a value of 0.5 for the trend damping parameter.

- **Optimise once**. Here the initial section of the time series is used, the optimal parameters of the smoothing methods were found at the end of the first period (the 23rd month) and were **not** re-optimized at the end of periods 31, 38, 45 and 53. This procedure was followed by Fildes (1992).

- **Optimise at every horizon**. In this case all the available time series information is used. The parameters were optimized at each ending period before the forecasts were being made. This approach, used earlier in the paper, is recommended by Chatfield (1978). It has the advantage that the optimal parameters calculated each time incorporate any new information. This approach permits a fair comparison between methods as the parameters of all other methods used are optimized each time forecasts are made. A compromise approach can also be envisioned which optimizes the smoothing parameters periodically rather than each time forecasts are needed. This analysis was performed on the telecommunications data whose series all have the same number of observations. The wide variation in series length of the M-Competition data would lead to varying numbers of time origins making a similar analysis more complex and this exercise was not attempted.

The results of these different approaches to setting the parameter values are shown in Table 9. For Simple Exponential Smoothing, optimising α once at month 23 produces a dramatic improvement in MAPE over the arbitrary choice of parameter value. Optimising at each time origin makes only a minor further improvement. However, it should be remembered that this method does not include a trend, thus the method simply tries to adjust the level as quickly as possible by choosing α equal to 1 or close to it, (i.e. it is not an appropriate method for this data set). The background to the suggestion of $\alpha = 0.1$ was a situation with no trend and a large random component and thus it should only be used when this is believed to be the case.

For Holt's Exponential Smoothing, the forecasts when optimizing at each time origin are better than optimizing only once. In turn, these results are better than arbitrarily setting the values for the parameters according to the early literature. This is particularly true for shorter forecasting horizons. Moreover, the third approach produces no "catastrophic" forecasts (or errors) as was the case when Fildes (1992) optimized the smoothing parameters only once. For Damped Exponential Smoothing, the differences are much more pronounced than those of Single or Holt's. The forecasts of optimizing at each ending period are superior to those of optimizing only once which in turn are superior for setting the parameter values to the arbitrary values suggested. Moreover, the differences between the three approaches become more pronounced with longer forecasting horizons.

Choosing to use fixed arbitrary values was a convenient expedient when computing facilities were less advanced and optimisation had a non-negligible cost. Instead of using the values suggested in the literature, an alternative approach is to use values suggested by the set of time series being forecast. For both Holt's and Damped methods, the most frequently occurring pair or triplet of parameter values found optimal using the first twenty three observations over all the telecommunications series was identified. These parameter values were used as fixed values for all series for the remaining time origins. For Holt's method, the most frequently occurring pair of values for level and trend smoothing parameters belonged to the intervals (0.8, 1.0) and (0, 0.2) respectively. Just over half the series fell into this classification, the mid-interval values of 0.9 and 0.1 were used. For Damped Smoothing, the most frequently occurring triplet of values for level, trend and damping the trend belonged to the intervals (0.75,1.0), (0.75, 1.0) and (0.25, 0.5) respectively. Forty three of the series fell in to this classification, the mid interval values used were 0.87, 0.87 and 0.37. These values for smoothing level and trend are in sharp contrast to those taken from the literature, showing a far greater tendency to adjust estimates of both level and trend. Although for this modal triplet, the value of the damping parameter is not extreme; for 18% of series this parameter is less than 0.25, showing that in these cases the method rapidly damps the trend as the horizon increases. The forecasting performance of this approach is also shown in table 9. These values based on information from all the series offer better performance than the literature based values. For Holt's method the performance is better than either of the optimising approaches. The same is true for Damped Smoothing apart from the 18 month horizon.

Table 9. The MAPEs for the Smoothing Methods under different regimes for the choice of parameter value using the telecommunications data (MAPEs averaged over last four time origins)

Method	Parameter setting regime	Forecasting Horizon			
		1	6	12	18
Single	Arbitrary	10.15	16.40	29.23	42.40
	Optimize Once	1.51	7.61	18.21	29.18
	Optimize Each Time	1.46	7.56	18.14	29.11
Holt's	Arbitrary	4.10	6.91	11.02	15.89
	Optimize Once	1.47	5.46	10.54	16.56
	Optimize Each Time	1.33	4.90	9.45	14.76
	Most common values	1.29	4.47	8.39	12.98
Damped	Arbitrary	10.43	17.19	31.59	45.85
	Optimize Once	1.54	6.56	15.76	25.33
	Optimize Each Time	1.37	5.60	12.60	19.48
	Most common values	1.34	5.11	12.44	20.75

5. Conclusions

This study has shown that the findings of the M-Competition hold for the telecommunications data. Firstly, complex or statistically sophisticated methods did not, in general, outperform simple ones. For example the ARIMA method was one of the least accurate for longer horizons. Secondly, the performance of various methods varied depending upon the specific accuracy measure used to evaluate the results. In addition, the importance of a comparison over multiple time origins was demonstrated. Thirdly, the combining of Holt's and Damped Smoothing outperformed the individual methods of Holt and Damped, in terms of MAPE, and did well overall in comparison to other methods and accuracy criteria. Finally, the performance of the various methods depended upon the length of the forecasting horizon involved.

This study reinforces the conclusions of the M-Competition. At a practical level, relatively accurate forecasts can be achieved with simple methods and through combining forecasts. The superior performance of ARARMA versus ARIMA models over longer horizons reinforces the finding that Parzen's method of identifying and extrapolating the trend in the data, if any, is more appropriate than Box-Jenkins' method of differencing the data to achieve stationarity in the mean (see Meade and Smith, 1985).

One of the most interesting findings of this study is the excellent performance of Robust Trend on the telecommunications data. By extending the range of methods included in Fildes (1992), we have confirmed that Robust Trend outperformed all other methods for practically all forecasting horizons and accuracy measures. This consistent superiority is directly related to the homogeneous structures of the time series used in this data set. Even with more heterogeneous structures of time series such as those of the M-Competition, Robust Trend performed as well as Holt's or Single Smoothing or ARIMA models.

The contrasting performance of Single, Holt and Damped Smoothing was examined when their parameters were set arbitrarily, only once at the first time origin, or were optimized each time forecasts were made. It was confirmed that the better approach was to re-optimize the smoothing parameters at each ending period. In addition to producing more accurate forecasts overall, this approach did not produce any very large errors. Many software designers need to modify their implementation of exponential smoothing accordingly. Fixed values for smoothing parameters recommended in the early literature were found to produce poor results. However, a cross-sectional analysis of the optimal parameters for time origin 23 suggested fixed values that performed better overall than repeated optimisation.

This study demonstrates that the performance ranking of forecasting methods depends on the sub-population of time series considered. If the structures of the time series are homogeneous, like those of the telecommunications data, then this property can and should be exploited. The benefit of choosing a specially designed method for the telecommunications data can be seen from the mean or median APEs where the value for Robust Trend is about half that of Damped smoothing. This substantial reduction in uncertainty is likely to lead to appreciable monetary savings.

REFERENCES

Armstrong, J.S. and F. Collopy, 1992, Error measures for generalizing about forecasting methods: empirical comparisons (with discussion), International Journal of Forecasting, **8**, 69-80, 99-111.

Brown, R.G., 1963, Smoothing, Forecasting and Prediction (Prentice Hall, New Jersey, USA).

Chatfield, C., 1978, The Holt Winters forecasting procedure, Applied Statistics, **27**, 264- 279.

Chatfield, C., 1992, A commentary on error measures, International Journal of Forecasting, **8**, 100-102.

Clemen, R., 1989, Combining forecasts: a review and annotated bibliography with discussion, International Journal of Forecasting, **5**, 559-608.

Fildes, R., 1992, The evaluation of extrapolative forecasting methods (with discussion), International Journal of Forecasting, **8**, 81-111.

Fildes, R. and S. Makridakis, 1995, The impact of empirical accuracy studies on time series analysis and forecasting, International Statistical Review, **63**, 289-308.

Gardner, E.S. Jr., and E. McKenzie, 1985, Forecasting trends in time series, Management Science, **31**, 1237-1246.

Geurts, M.D. and J.P. Kelly, 1986, Forecasting retail sales using alternative models, International Journal of Forecasting, **2**, 261-272.

Grambsch, P., and W.A. Stahel, 1990, Forecasting demand for special services, International Journal of Forecasting, **6**, 53-64.

Hill, G. and R. Fildes 1984, The accuracy of extrapolation methods; an automatic Box-Jenkins package Sift, Journal of Forecasting, **3**, 319-323.

Coutie, G.A., O.L.Davies, C.H. Hossell, D.W.G.P. Millar and A.J.H. Morrell, 1964, Short Term Forecasting, ICI Monograph No 2, Oliver and Boyd, Edinburgh, UK.

Koehler, A.B., and E.S. Murphree, 1988, A comparison of results from state space forecasting with forecasts from the Makridakis competition, International Journal of Forecasting, **4**, 45-55.

Ledolter, J., 1989, The effect of additive outliers on forecasts from ARIMA models, International Journal of Forecasting, **5**, 231-240.

Libert, G., 1984, The M-Competition with a fully automatic Box-Jenkins procedure, Journal of Forecasting, **3**, 325-328.

Makridakis, S., 1986, The art and science of forecasting: an assessment and future directions, International Journal of Forecasting, **2**, 15-40.

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

Makridakis, S., et al., 1984, The forecasting accuracy of major time series methods, (Wiley, New York).

Makridakis, S., et al., 1993, The M-2 Competition: A real-time judgmentally based forecasting study, International Journal of Forecasting, **9**, 5-23.

Makridakis, S., and R. Winkler, 1983, Average of forecasts: some empirical results, Management Science, **29**, 987-996.

Meade N and I. Smith, 1985, ARARMA vs ARIMA - a study of the benefits of a new approach to forecasting, Omega, **13**, 519 - 534.

Parzen E., 1982, ARARMA models for time series analysis and forecasting, Journal of Forecasting , **1**, 67-82.

Simmons, L.F., 1986, M-Competition - A closer look at NAIVE2 and median APE: a note, International Journal of Forecasting, **4**, 457-460.

Stellwagen, E.A., and R.L. Goodrich, 1991, Forecast Pro - Batch Edition, (Business Forecast Systems, Inc., Belmont, MA).

Tashman, L.J. and J.M. Kruk, 1996, The use of protocols to select exponential smoothing procedures: a reconsideration of forecasting competitions. International Journal of Forecasting, **12**, 235 - 253.

Wu, L.S., J.R.M. Hosking and N. Ravishanker, 1993, Reallocation outliers in time series, Applied Statistics, **42**, 301-313.

Appendix 1. The Robust Trend model

For a time series, X_t, the model is

$$Z_t = X_t - X_{t-1} = \mu + \sigma\, e_t$$

where e_t is i.i.d. stable random variable. At time T, the k step ahead forecast is

$$X_{T+k} = X_T + k\,\hat{\mu}_T$$

The robust estimate of the trend is

$$\hat{\mu}_T = M_T + \frac{m_T}{T} \sum_{t=1}^{T} \psi\left(\frac{Z_t - M_T}{m_T}\right)$$

where M_T is the median of $(Z_1, \ldots, Z_T)$ and m_T is the median of $(|Z_1 - M_T|, \ldots, |Z_T - M_T|)$. The robustness of the method lies in its treatment of outliers, the response function ψ is of a "three part re-descending" type:

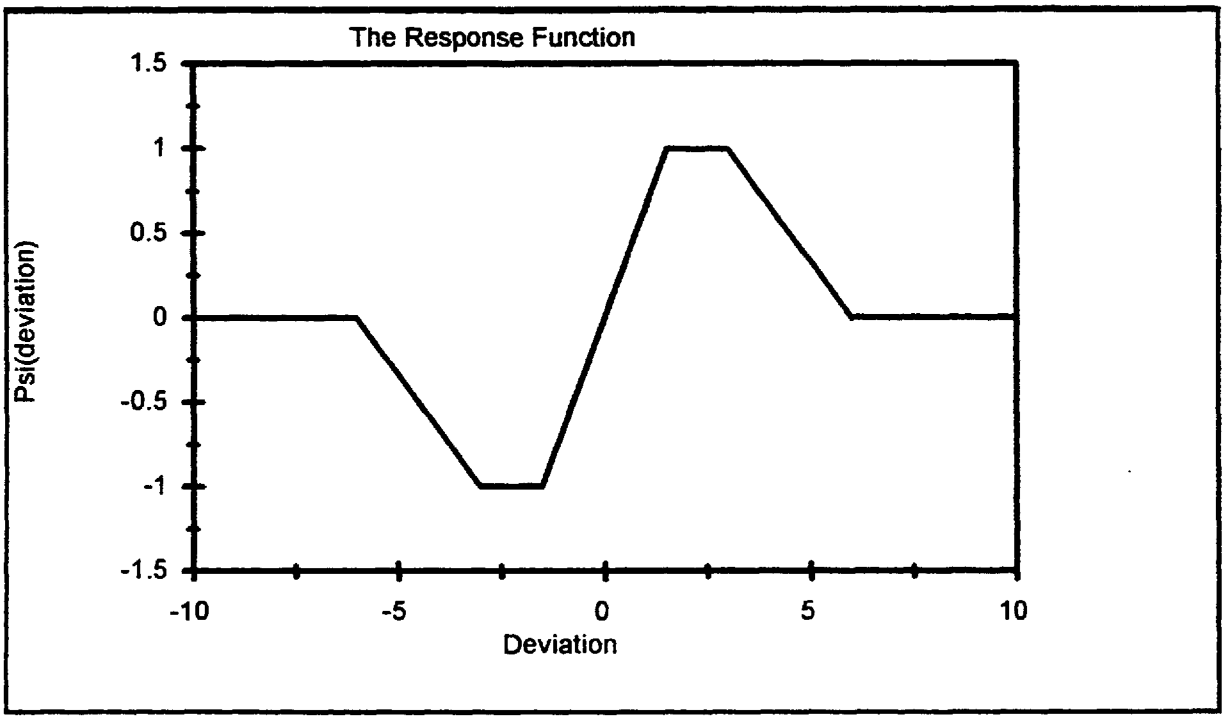
$$\psi(x) = sign(x)\ \max\left[\ \min(\ |2x/3|,\ 1.0,\ 2 - |x/3|\ ),\ 0\ \right]$$

The response is shown in Figure A1.

As can be seen from the figure, the effect of very large deviations are not allowed to affect the trend adjustment.

For fuller details about this method, especially the calculations of robust prediction intervals, see Grambsch and Stahel (1990).

Figure A1. The response function ψ



Biographies

Robert FILDES is Editor-in-Chief of the International Journal of Forecasting and Professor in the Management School, Lancaster University. He has published four books in forecasting and planning. He was co-founder of the International Institute of Forecasters and the Journal of Forecasting in 1982 and the International Journal of Forecasting in 1985. He has published numerous articles in academic journals, including Management Science, Economica and the Journal of the Operational Research Society. In 1985 he was a Visiting Fellow at Bell Communications Research, U.S.A. He is currently Vice-President of the UK Operational Research Society.

Michèle HIBON is currently a Research Associate at INSEAD. For the last several years she has been working on various studies dealing with forecasting accuracy of time series methods and more recently in the area of Group Decision Support Systems. She graduated from the University of Paris, holds a degree in Science and a diploma in advanced studies in physics.

Spyros MAKRIDAKIS is a Research Professor at INSEAD, Fontainbleau, France. His Ph.D. is from New York University. He has consulted worldwide in the area of forecasting and has held teaching positions with several European and American institutions; as a research fellow at IIM in Berlin, an ICAME fellow at Stanford University and a Visiting Scholar at MIT and Harvard. Professor Makridakis is the author of approximately 100 articles and papers in various publications and has also authored or co-authored eighteen books including *Forecasting, Planning and Strategy for the 21st Century* (Free Press, 1990).

Nigel MEADE is Reader in Management Science at the Management School, Imperial College, University of London. His research interests are statistical model building in general and applied time series analysis and forecasting in particular. He is currently Chairman of the UK OR Society Forecasting Study Group.