M3 – Competition

By Michèle Hibon* and Spyros Makridakis**

99/70/TM

* Senior Research Fellow at INSEAD,
Boulevard de Constance, 77305 Fontainebleau Cedex, FRANCE
E-mail : Michele.Hibon@insead.fr

** Research Professor of Decision Sciences at INSEAD,
Boulevard de Constance, 77305 Fontainebleau Cedex, FRANCE

M3 – Competition

ABSTRACT

The major aims of the M3-Competition are to extend and replicate the findings of the M- and M2-ones. The extension involves the inclusion of more researchers, more methods (in particular in the area of neural networks and expert systems) and most importantly more series as the database of the M3-Competition has been enlarged to include 3003 time series. In terms of replication our purpose is to determine if the four major conclusions of the M- and M2-Competitions:

(1) *Statistically sophisticated or complex methods do not necessarily produce more accurate forecasts than simpler ones.*

(2) *The rankings of the performance of the various methods vary according to the accuracy measure being used.*

(3) The accuracy of the combination of various methods outperforms, on average, the Individual methods being combined and does well in comparison with other methods;

(4) The performance of the various methods depends upon the length of the forecasting horizon.

Still apply.

Keywords: Forecasting competition, M-Competition, Forecasting accuracy

M3 – Competition

This study has been done on the enlarged, new database of **3003 series** and includes **24 methods**.

The time series have been selected on a quota basis:

- 6 different types of series: micro, industry, finance, demographic and other,
- 4 different time intervals between successive observations (yearly, quarterly, monthly and other).

The historical values of each series are

At least 14 observations for yearly data,

At least 16 observations for quarterly data,

At least 48 observations for monthly data,

At least 60 observations for other data,

The time horizons of forecasting are:

6 periods for yearly data,

8 periods for quarterly data,

18 periods for monthly data,

8 periods for other data.

The **3003 time series** are distributed as follows

Time interval between successive observations	**TYPES OF TIME SERIES DATA**						
	Micro	Industry	Macro	Finance	Demographic	Other	TOTAL
Yearly	**146**	**102**	**83**	**58**	**245**	**11**	**645**
Quarterly	**204**	**83**	**336**	**76**	**57**		**756**
Monthly	**474**	**334**	**312**	**145**	**111**	**52**	**1428**
Other	**4**			**29**		**141**	**174**
TOTAL	**828**	**519**	**731**	**308**	**413**	**204**	**3003**

Methods	Competitors	Description
Naïve/Simple		
1. NAÏVE 2	M. Hibon	Deseasonalized Naïve (Random Walk)
2. SINGLE	M. Hibon	Single Exponential Smoothing
ExplicitTrendModels		
3. HOLT	M. Hibon	Automatic Holt's Linear Exponential Smoothing (2 parameter model)
4. ROBUST-TREND	N. Meade	Non parametric version of Holt's linear model with median based estimate of trend
5. WINTER	M. Hibon	Holt-Winter's linear and seasonal exponential smoothing (2 or 3 parameter model)
6. DAMPEN	M. Hibon	Dampen Trend Exponential Smoothing
7. PP autocast	H. Levenbach	Damped Trend Exponential Smoothing
8. THETA-sm	V.Assimakopoulos	Successive smoothing plus a set of rules for dampening the trend
9. COMB S/H/D	M. Hibon	Combining 3 Methods : Single/Holt/ Dampen
Decomposition		
10. THETA	V.Assimakopoulos	Specific decomposition technique , projection and combination of the individual components
ARIMA/ARARMA Model		
11. BJ-automatic	M. Hibon	Box Jenkins methodology of "Business Forecast System"
12. AUTOBOX 1 13. AUTOBOX 2 14. AUTOBOX 3	D. Reilly	Robust ARIMA univariate Box-Jenkins with/without Intervention Detection
15. AAM 1 16. AAM 2	G. Melard, J. M. Pasteels	Automatic ARIMA modelling with/without intervention analysis
17. ARARMA	N. Meade	Automated Parzen's methodology with Auto regressive filter
Expert System		
18. ForecastPRO	R. Goodrich, E. Stellwagen	Selects from among several methods: Exponential Smoothing/Box Jenkins/Poisson and negative binomial models/Croston's Method/Simple Moving Average
19. SMARTFCs	C. Smart	Automatic Forecasting Expert System which conducts a forecasting tournament among 4 exponential smoothing and 2 moving average methods
20. RBF	M. Adya, S. Armstrong, F. Collopy, M. Kennedy	Rule-based forecasting: using 3 methods - random walk, linear regression and Holt's to estimate level and trend, involving corrections, simplification, automatic feature identification and recalibration
21. FLORES-PEARCE1 22. FLORES-PEARCE2	B.Flores, S. Pearce	Expert system that chooses among 4 methods based on the characteristics of the data
23. ForecastX	J. Galt	Runs tests for seasonality and outliers and selects from among several methods : Exponential Smoothing, Box-Jenkins and Croston's method
Neural Networks		
24. Automat ANN	K. Ord, S. Balkin	Automated Artificial Neural Networks for forecasting purposes

METHODS

The different methods have been classified in the following categories

- Naïve, simple methods
- Explicit Trend Models
- Decomposition
- ARIMA / ARARMA models
- Expert Systems
- Neural Networks

The array displayed on the previous page gives a list of the 24 methods that have been used in the competition with the name of the competitors and a short description.
In the appendix one can find a more detailed description of the new methods given by their authors.

ACCURACY MEASURES

The accuracy measures to describe the results of the competition reported in this paper are:

- MAPE: Symmetric Mean Absolute Percentage Error,

  If X is the real value and F is the forecast , the formula for the symmetric MAPE is :

$$\frac{|X - F|}{\frac{X + F}{2}} * 100$$

  By taking the symmetric MAPE, we avoid the problem of distortion we had with the regular MAPE if the actual values are close to zero.

- Ranking: Average Ranking,

  This is the average ranking of the symmetric absolute percentage error from each method for each horizon.

- Median APE: Median Absolute Percentage Error,
- Median RAE: Median Relative Absolute Error,
- RMSE: Root Mean Square Error.

THE RESULTS

We have calculated an overall average of the accuracy measures, but we also focus, on this paper, on a breakdown of these measures for each category and time interval between successive observations.

The figures 1 to 20 are the graphics of the average MAPE of yearly, quarterly, monthly and other series, in overall and per category. They allow comparing the performance of the methods that give the best results.

The tables 1 to 11 show the methods that give the best results, as follow:

-Tables 1, 2, 3, and 4: comparison of the 4 accuracy measures, on each category, for each time interval.

-Tables 5, 6, 7, 8, and 9: detailed results per category and per time interval for each accuracy measures.

-Table 10: comparison of the results given by MAPE on monthly data per category for short, medium and long step horizons.

-Table 11: comparison of the results over seasonal versus non-seasonal data.

The best way to understand the results is to consult the various tables carefully.

The different accuracy measures

Tables 1, 2, 3, 4 give the results for the four different accuracy measures that have been used. We can see that most of the time each accuracy measure identifies the same methods that give the best results for the different types of data.

Effects of the type of series

The table 5 shows for each category and each type of data, which methods do significantly better than others.

We found that THETA is performing very well for almost all types of data. Whereas other methods are more appropriate for a type/category of data:

ForecastPro for monthly data, for micro and industry data, ForcX for yearly data, for industry and demographic data,

RBF for yearly data, for macro data,

Robust-Trend for yearly data and for macro data,

Autobox2 for yearly and other data,

AAM1/AAM2 for finance data,

COMB-SHD for quarterly data,

ARARMA for other data and macro data,

If we consider the series as seasonal versus non- seasonal data, in overall average, ForecastPro is significantly better than any other methods for seasonal data and THETA for non-seasonal data

In overall average, THETA and ForecastPro are significantly better than all the other methods.

Effects of forecasting horizons

The results that are displayed in the different tables are averages over the different step horizons i.e. 1 to 6 for yearly data, 1 to 8 for quarterly data, 1 to 18 for monthly data and 1 to 8 for other data. A question, which might be of interest, is what would be the results if we consider the averages over short, medium and long term horizons.

The table 9 shows this result for monthly data assuming that:

Short term = average 1 to 3

Medium term = average 4 to 12

Long term = average 13 to 18

We found that the methods THETA and ForecastPro which are doing the best as overall, are also doing well when we consider separately short, medium and long term.

For short term, in addition to THETA and ForecastPro, there are SMARTFcs, AutomatANN and ForcX.

For medium term, in addition to ForecastPro and THETA there is ForcX.

For long term, in addition to THETA and ForecastPro there is RBF which is always doing better (for any kind of data) for long term horizon than for short term.

The combining of forecasts

The COMB-SHD method is a simple combination of the forecasts given by the three exponential smoothing methods: Single, Holt-Winter's and Dampened Trend. It gives good results especially for quarterly data (for each category) and for industry data (for yearly and quarterly data).

Complexity of the methods

THETA, which can be considered as a simple method, gives the best results for almost each type of data.

Flores-Pearce methods and RBF are methods, which are much time consuming, and they didn't produce more accurate results; and the neural-network method Automat ANN, didn't out-perform any other methods.

CONCLUSION

Comparison with the M-Competition

1. The performance of the various methods depends upon
   - The length of the forecasting horizon
   - The type (yearly, quarterly, monthly, others) of data
   - The category (micro, industry, macro, finance, demographic, other) of data.
2. Accuracy measures are consistent in the M3 Competition
3. The combination of the 3 exponential smoothing methods does better than the individual methods being combined and very well in comparison with the other methods
4. Statistically sophisticated or complex methods do not necessarily produce more accurate forecasts than simpler ones

New methods

Some specific new methods not used in the M- Competition perform consistently better than the others in specific circumstances:

THETA, ForecastPRO for **Monthly data**
THETA for **Quarterly data**
RBF, ForcX, THETA, Robust-Trend, Autobox2 for **Yearly data**
Autobox2, ARARMA, THETA, ForcX for **Other Data**

THETA, ForecastPRO for **Micro data**
ForecastPRO, Forcx, THETA for **Industry data**
RBF, ARARMA, THETA, Robust-Trend for **Macro data**
AAM1, AAM2 for **Finance Data**
ForcX for **Demographic Data**

ForecastPRO for **Seasonal Data**
THETA for **Non-Seasonal Data**

The performance of the different methods does not significantly differ for **short, medium and long** term

Who has won the competition?

It is not an appropriate question, and there is not a specific answer. It is more relevant to identify which methods are doing better than others are, for each specific type/category of data.

Figure 1 - Average Symmetric MAPE : Yearly Data

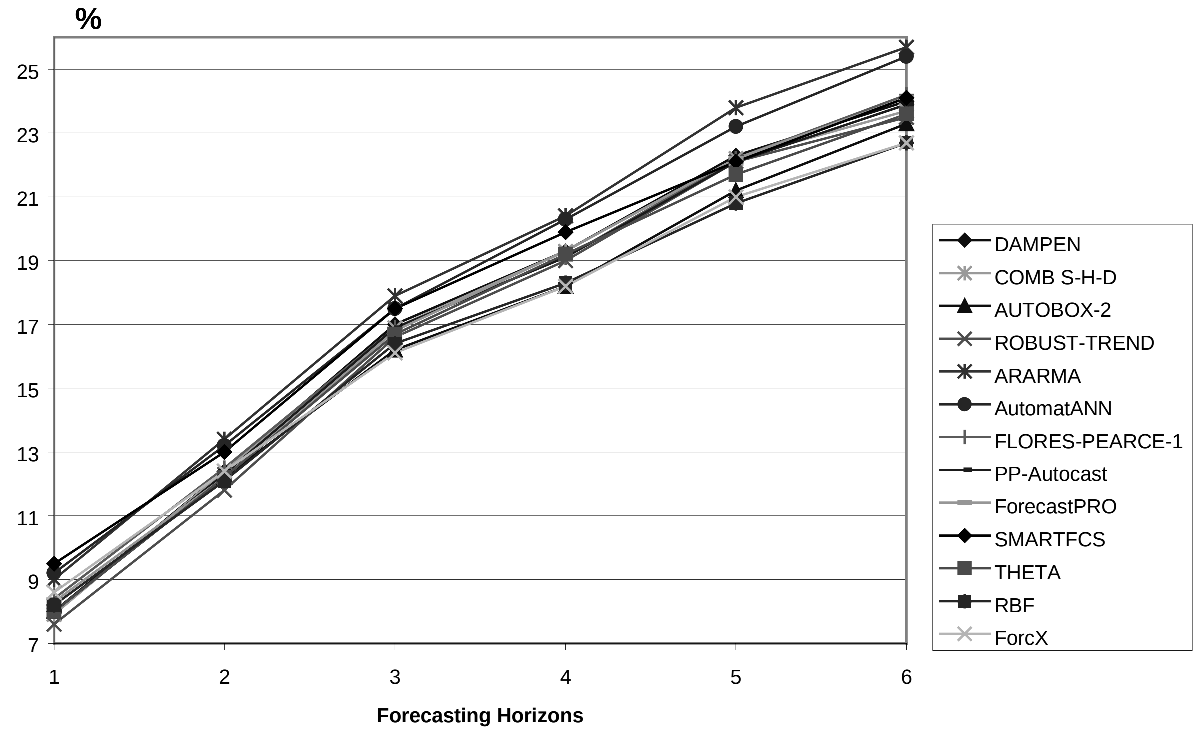


Figure 2 - Average Symmetric MAPE : Quarterly Data

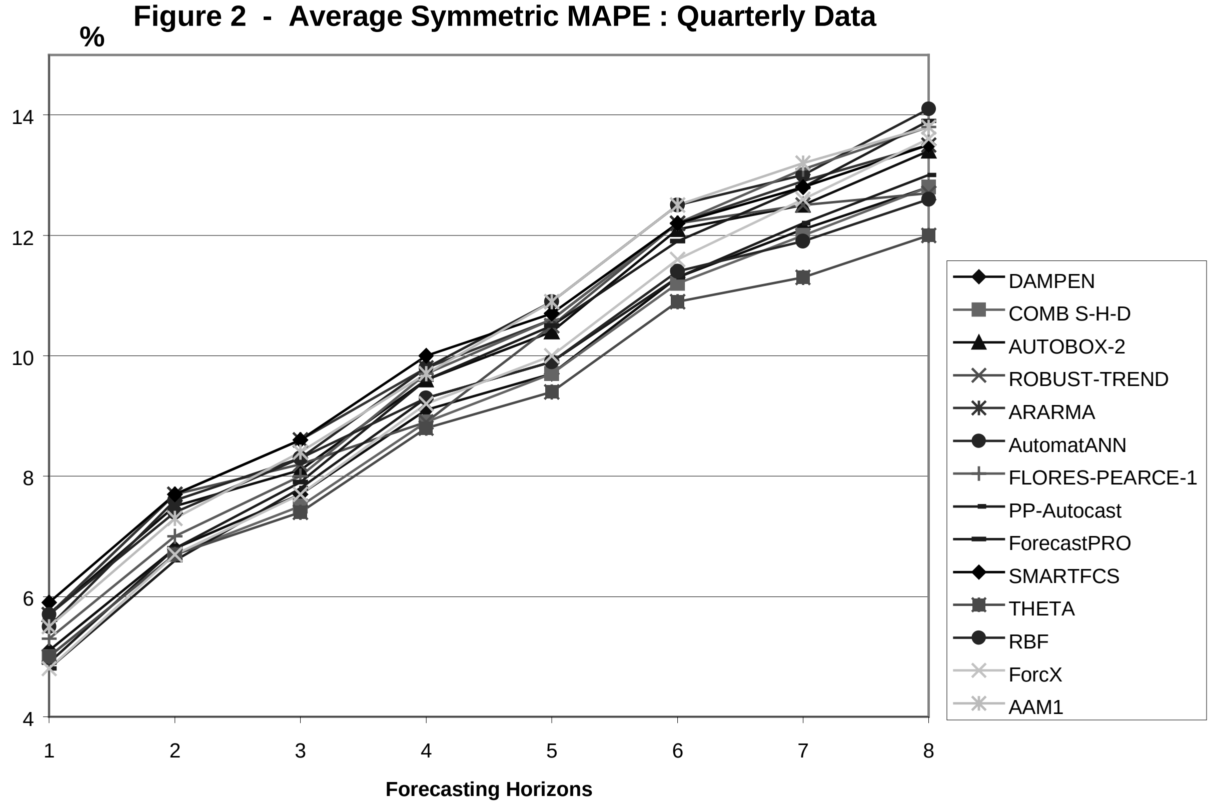


Figure 3 - Average Symmetric MAPE : Monthly Data

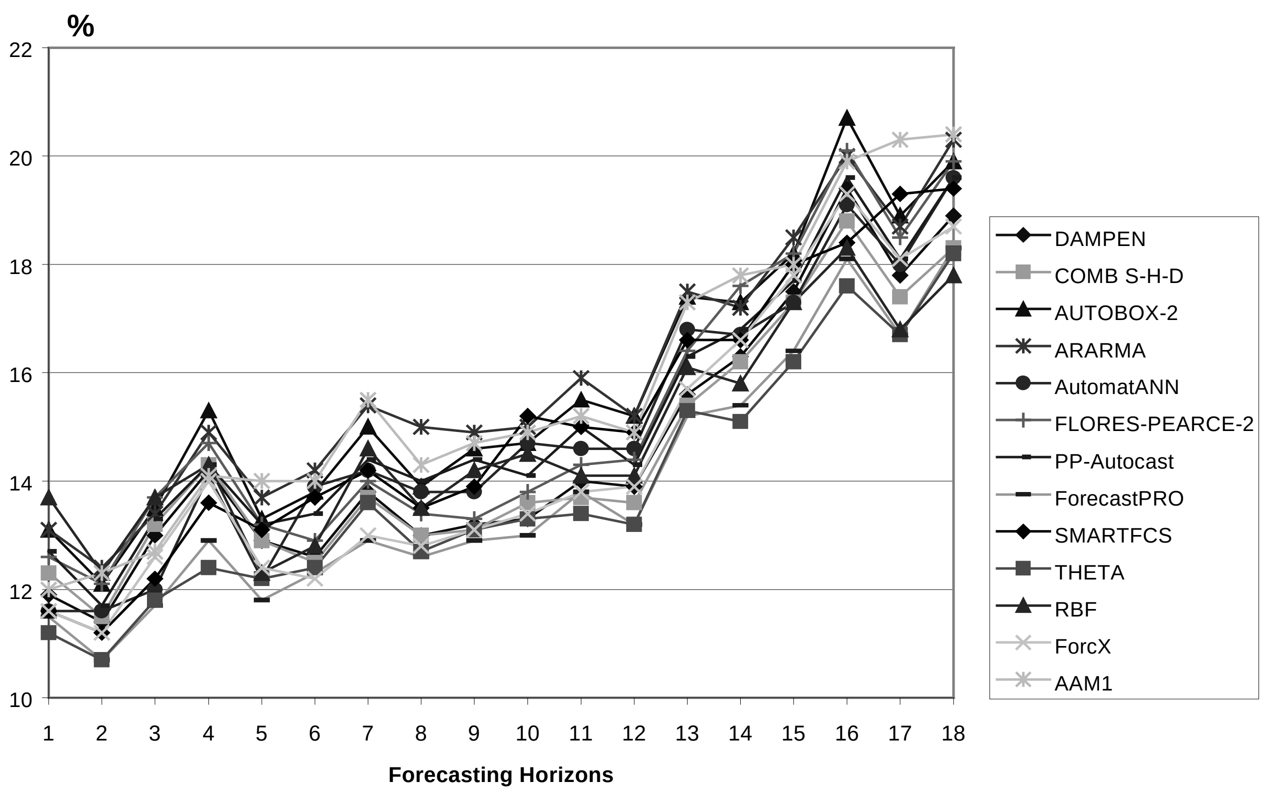


Figure 4 - Average Symmetric MAPE : Other Data

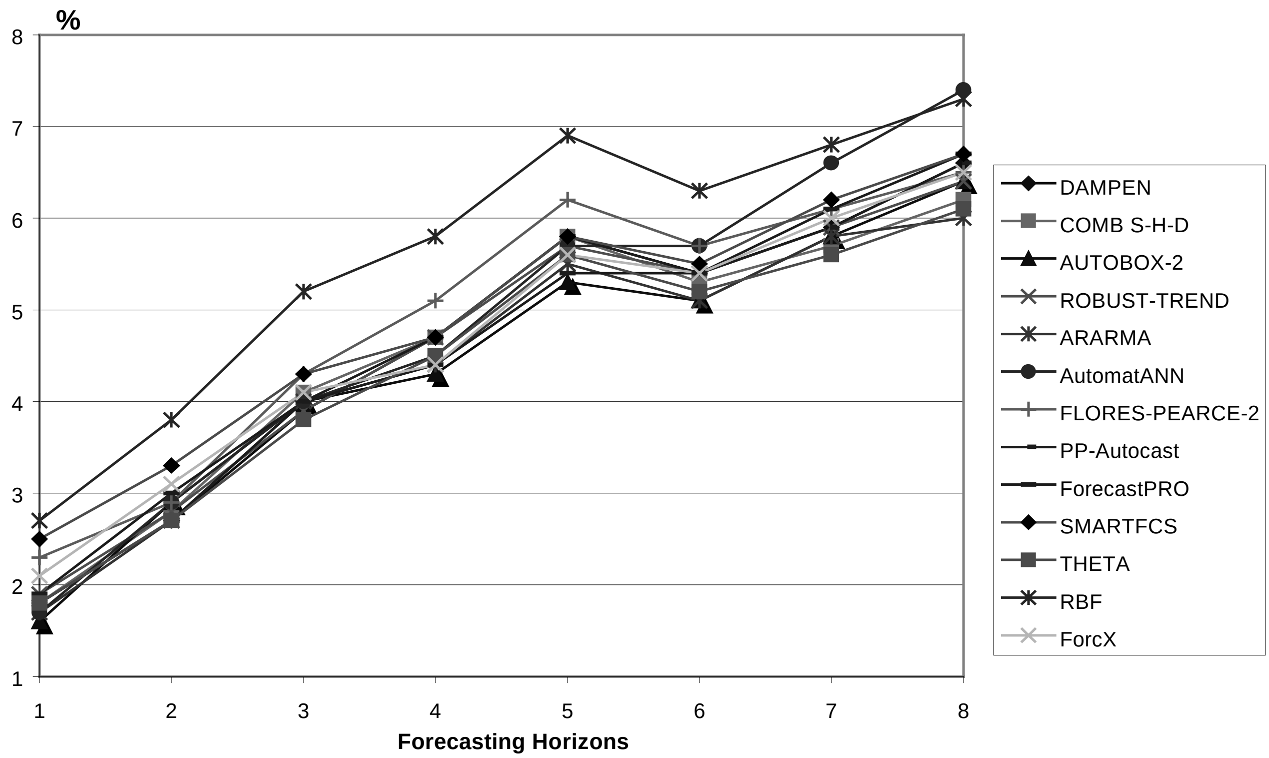


Figure 5 - Symmetric MAPE : Yearly - MICRO Data

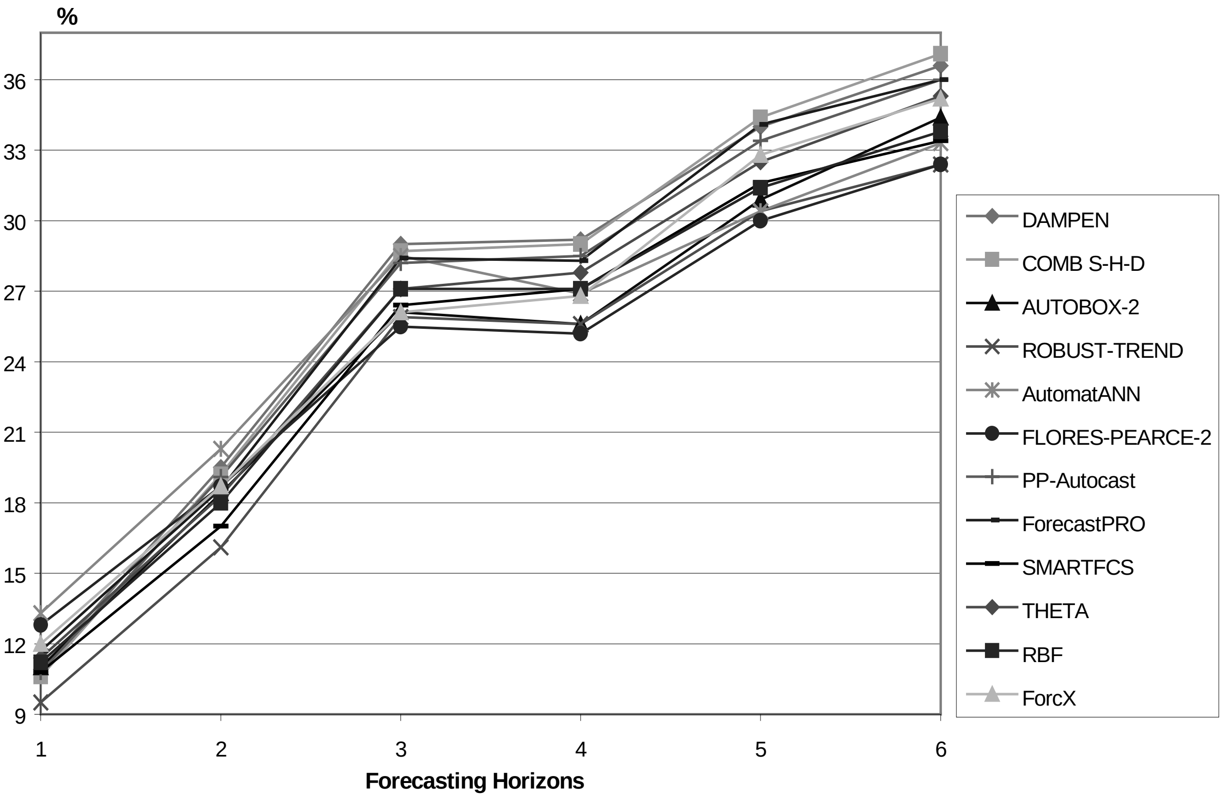


Figure 6 - Symmetric MAPE : Yearly - INDUSTRY Data

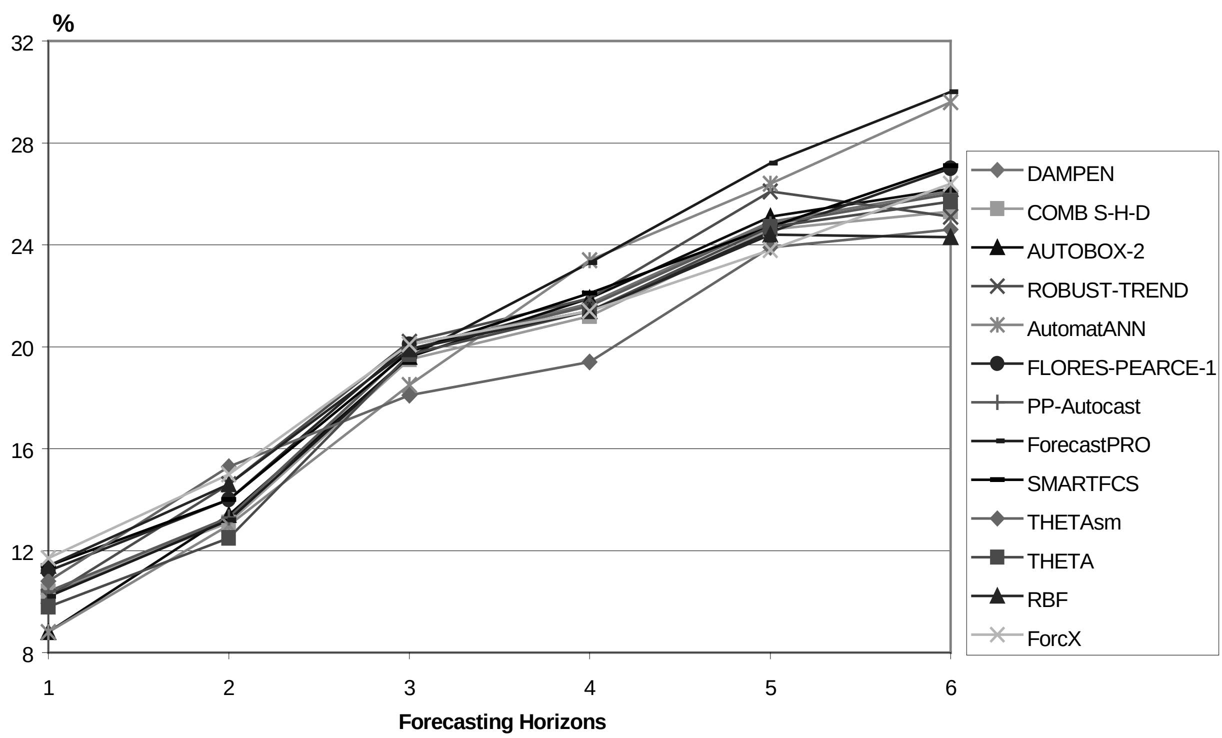


Figure 7 - Average Symmetric MAPE : Yearly - MACRO Data

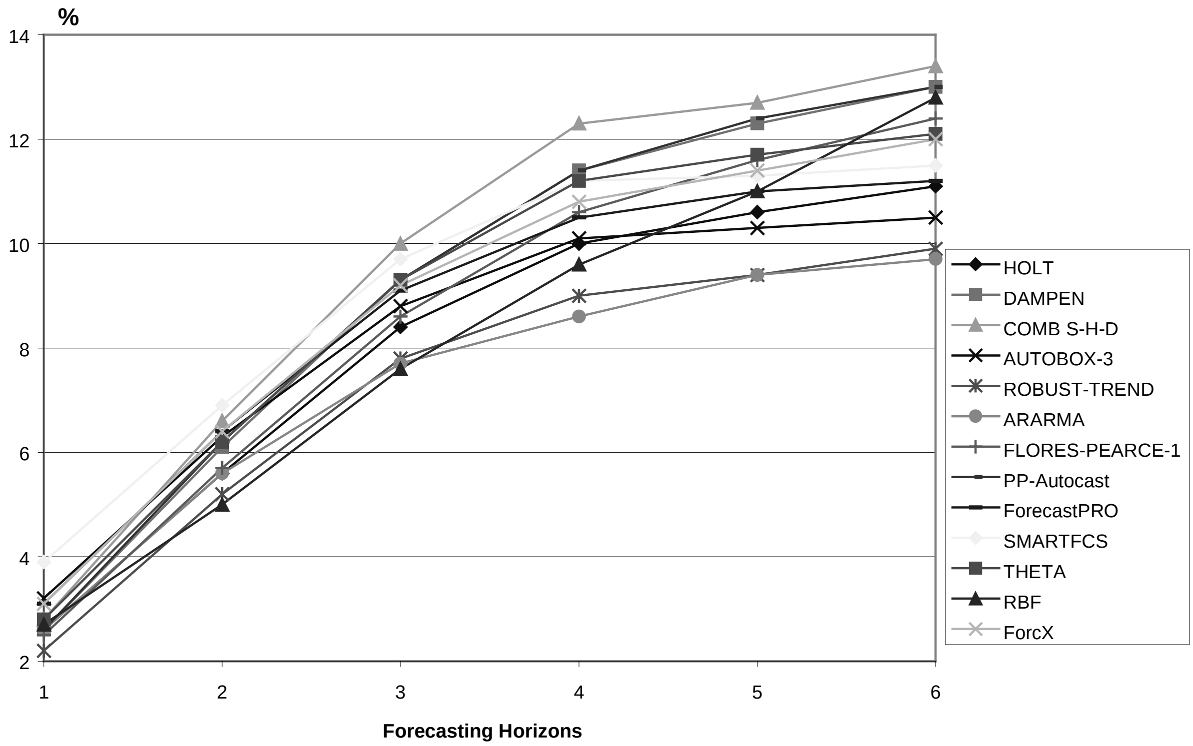


Figure 8 - Symmetric MAPE : Yearly - FINANCE Data

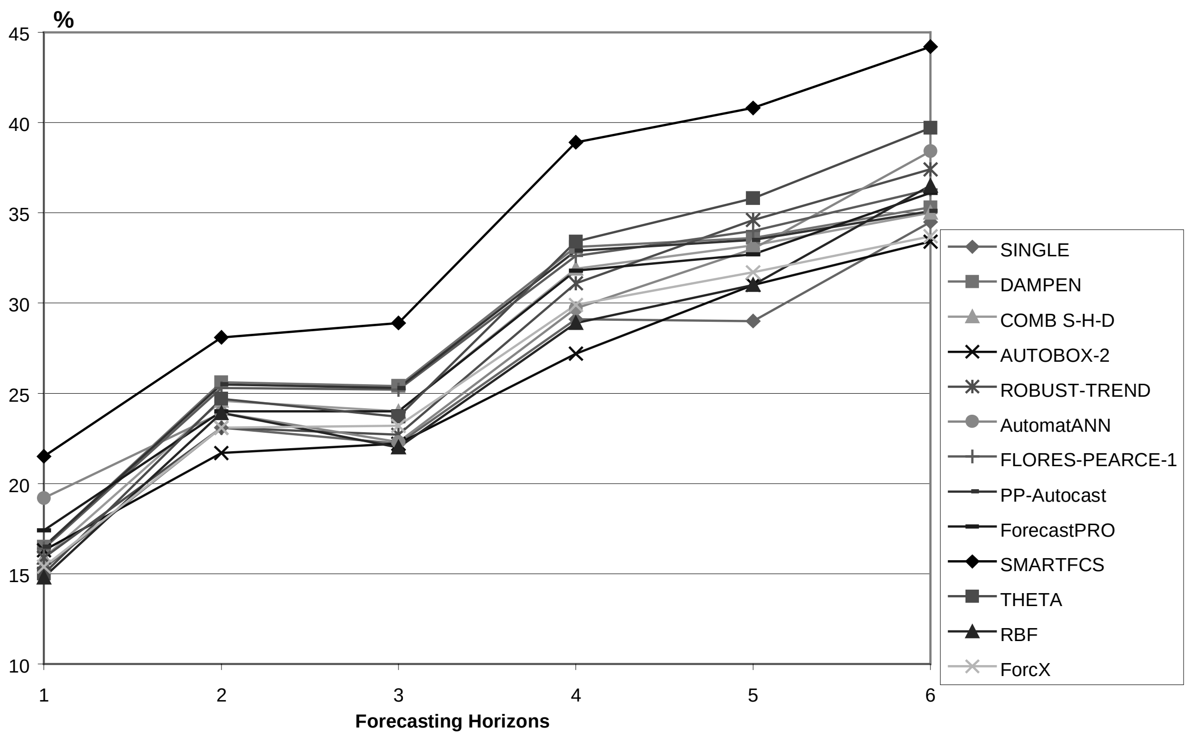


Figure 9 - Average Symmetric MAPE : Yearly - DEMOGRAPHIC Data

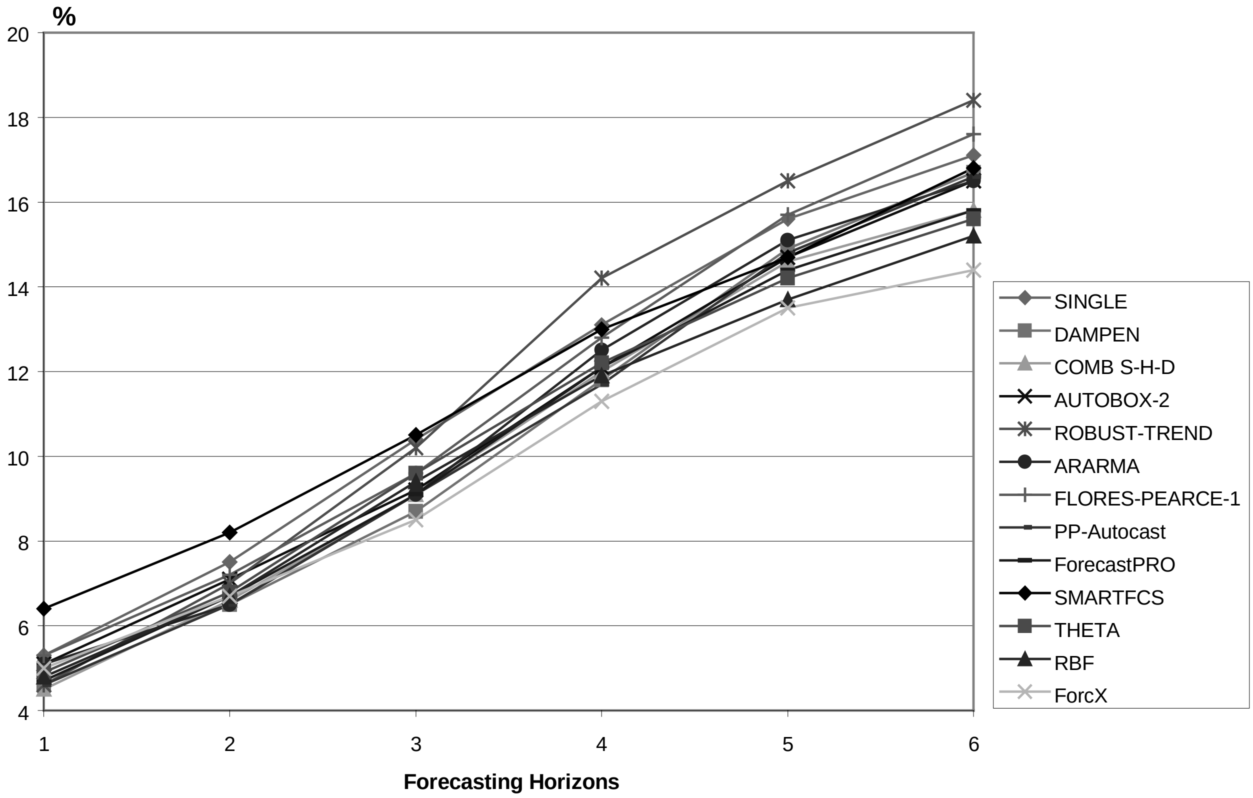


Figure 10 - Symmetric MAPE : Quarterly - MICRO data

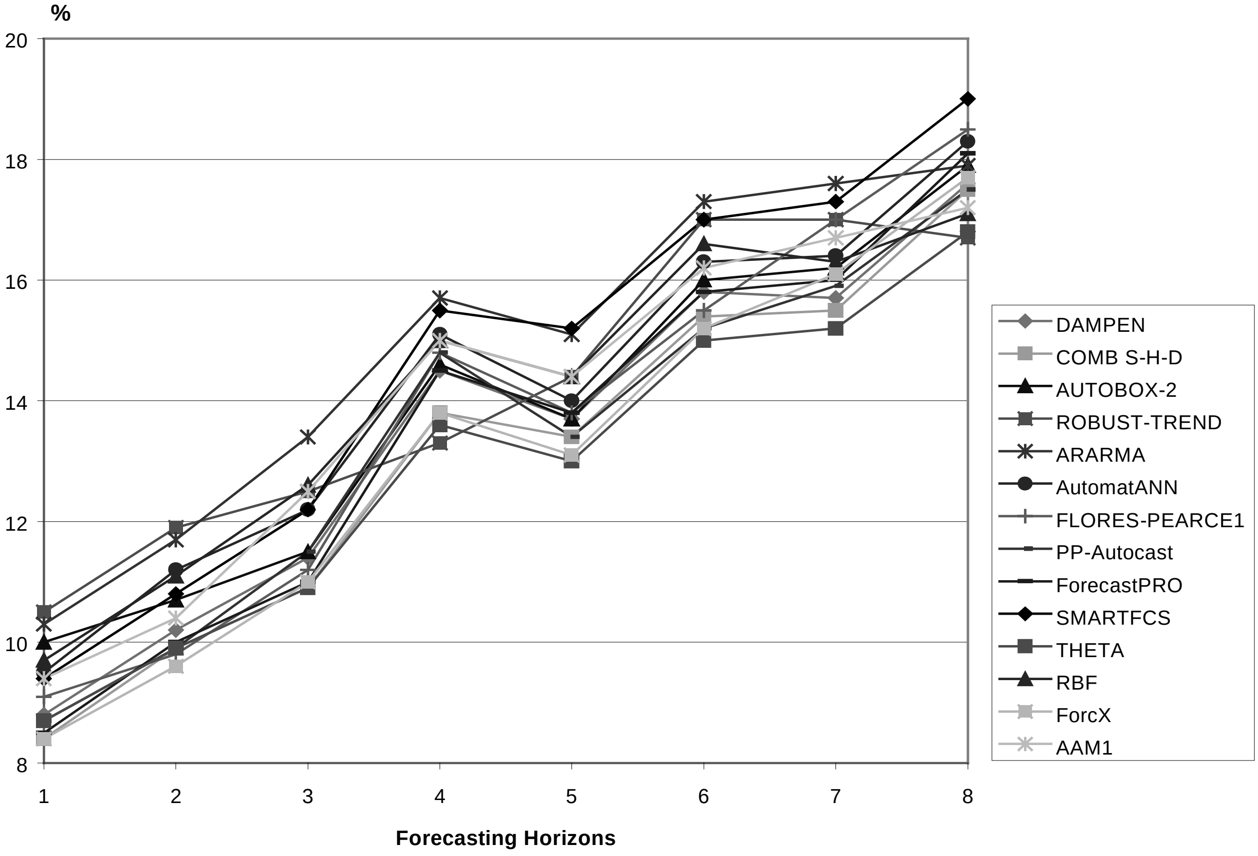


Figure 11 - Symmetric MAPE : Quarterly - INDUSTRY Data

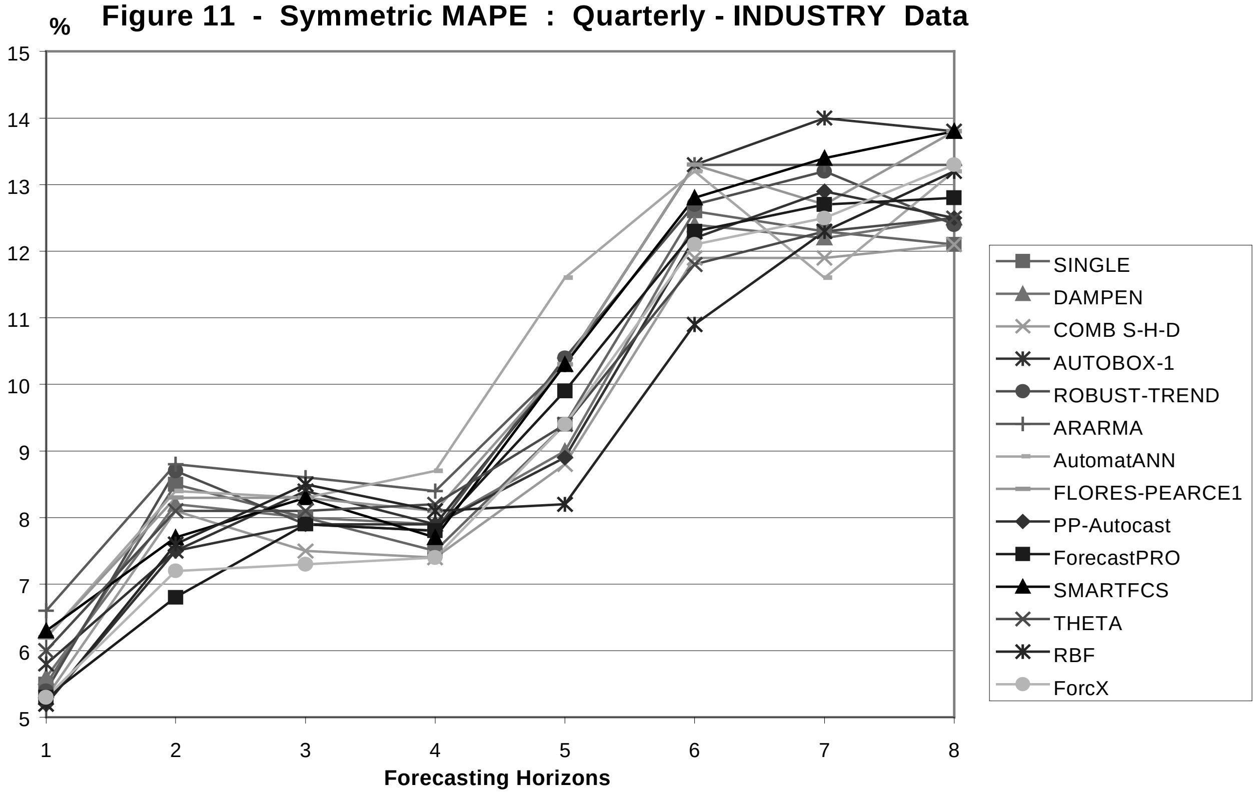


Figure 12 - Symmetric MAPE : Quarterly - MACRO Data

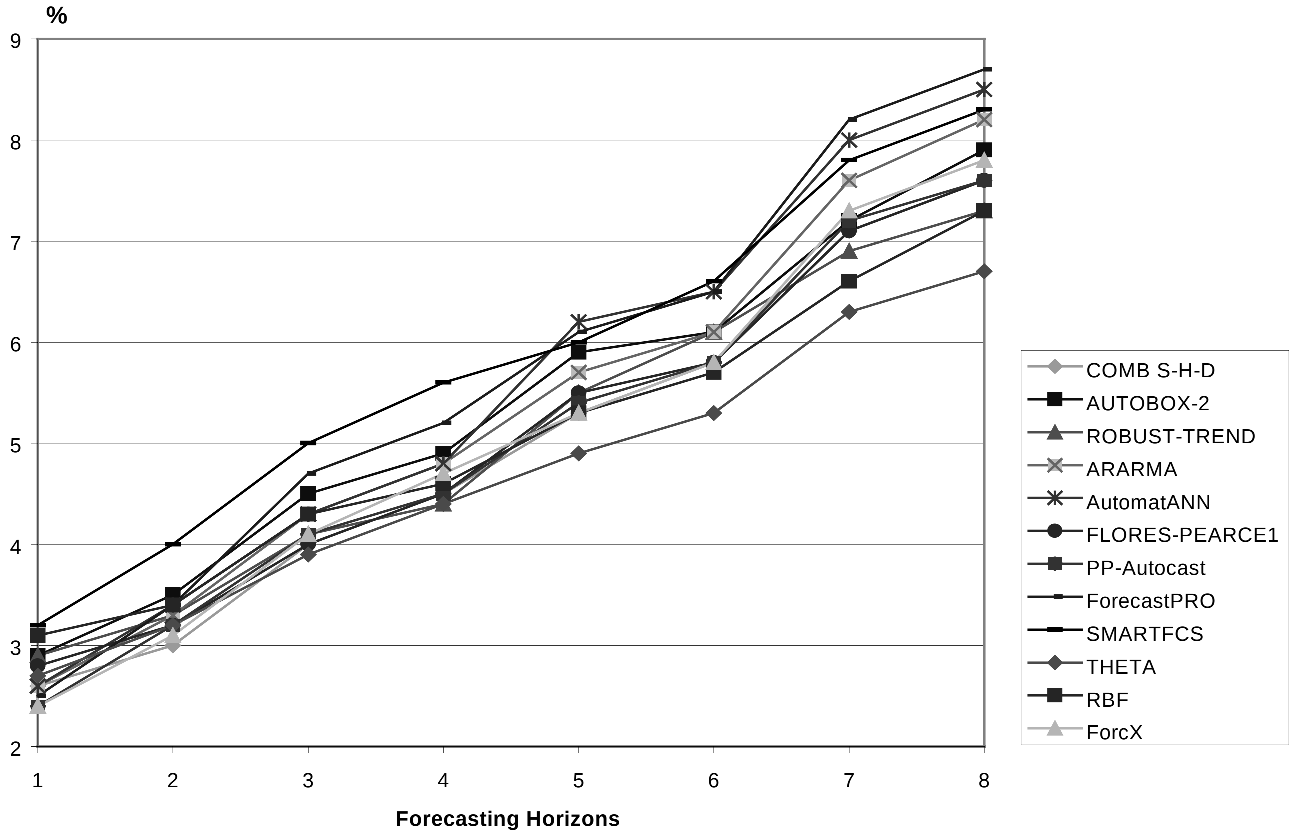


Figure 13 - Symmetric MAPE : Quarterly - FINANCE Data

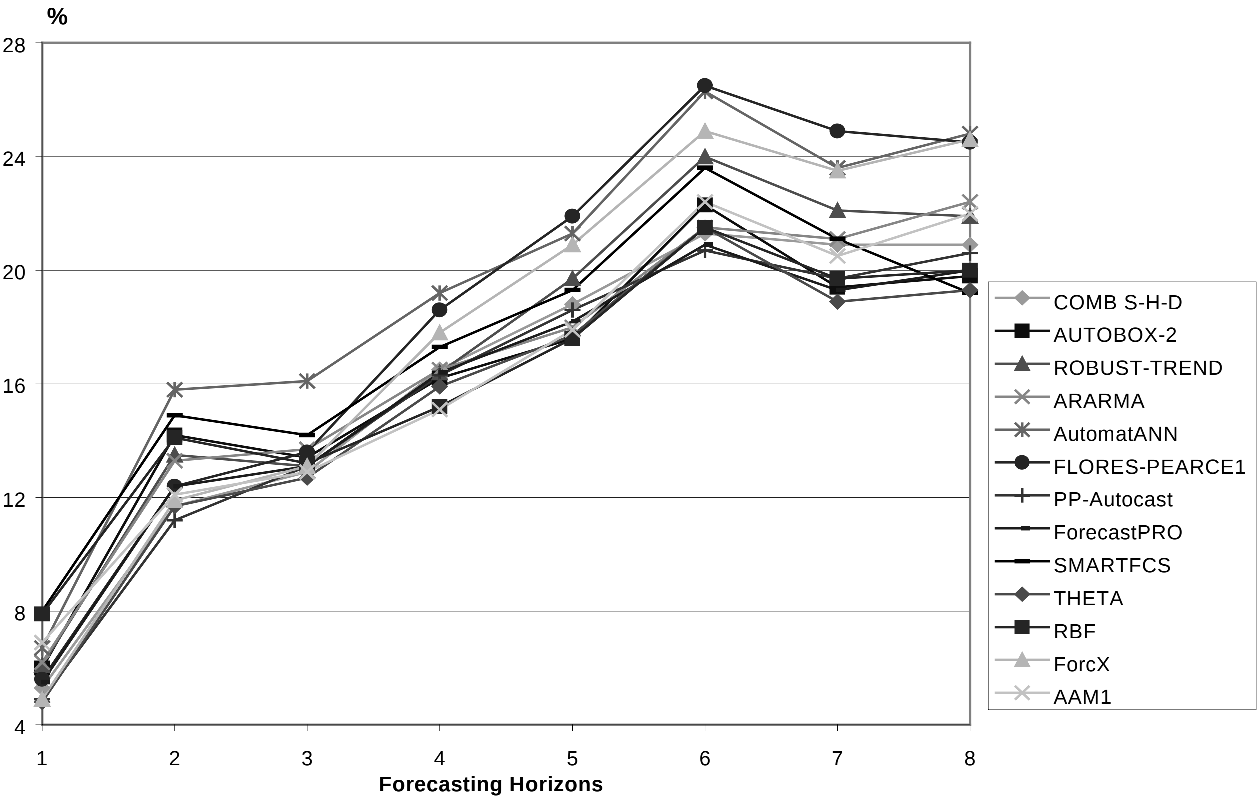


Figure 14 - Symmetric MAPE : Quarterly - DEMOGRAPHIC Data

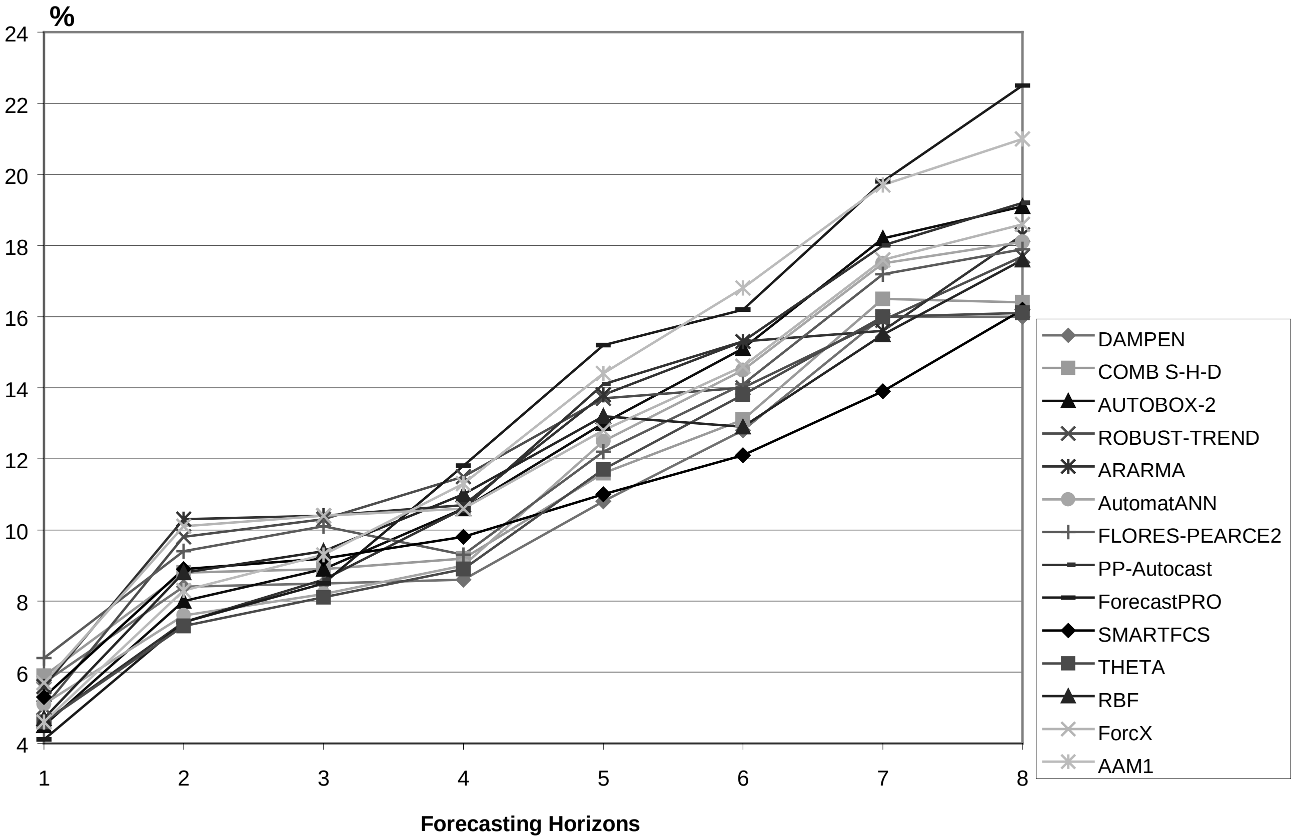


Figure 15 - Symmetric MAPE : Monthly - MICRO data

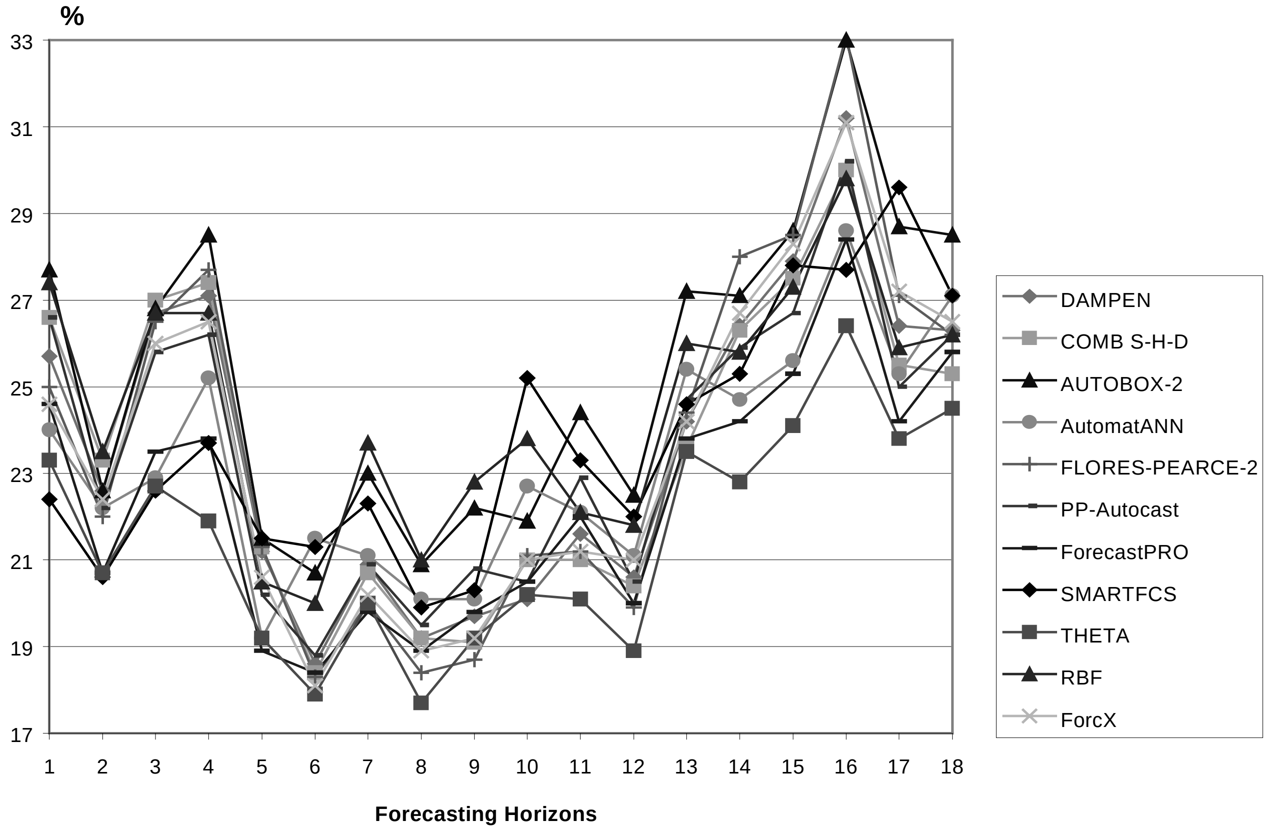


Figure 16 - Symmetric MAPE : Monthly - INDUSTRY data

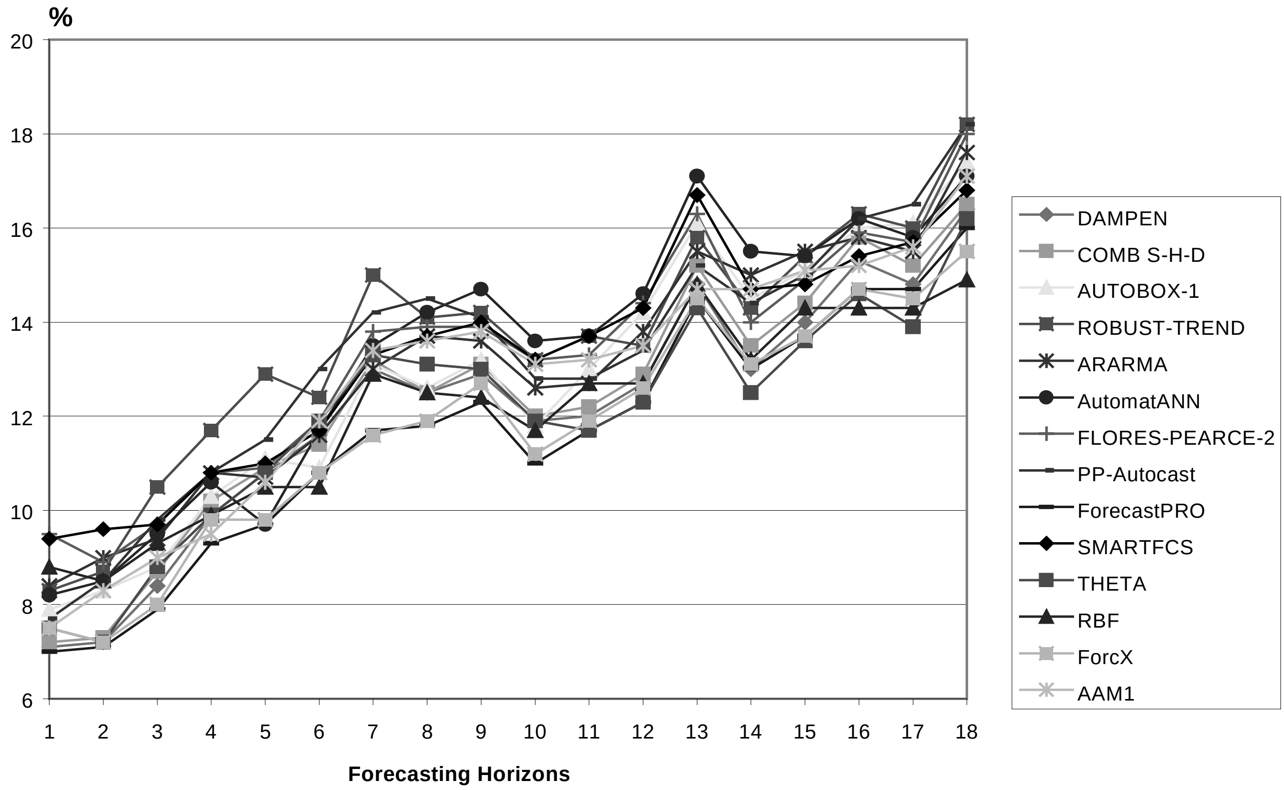


Figure 17 - Symmetric MAPE : Monthly - MACRO data

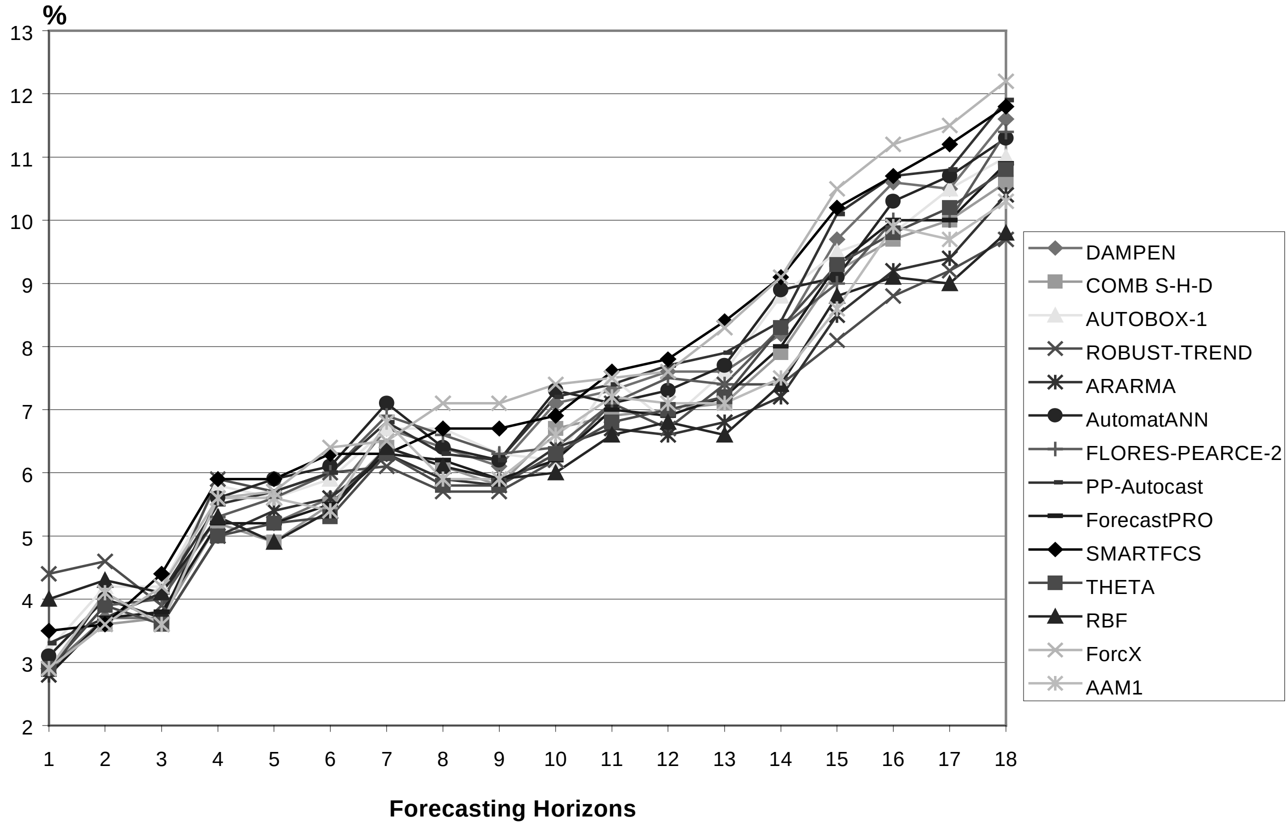


Figure 18 - Symmetric MAPE : Monthly - FINANCE data

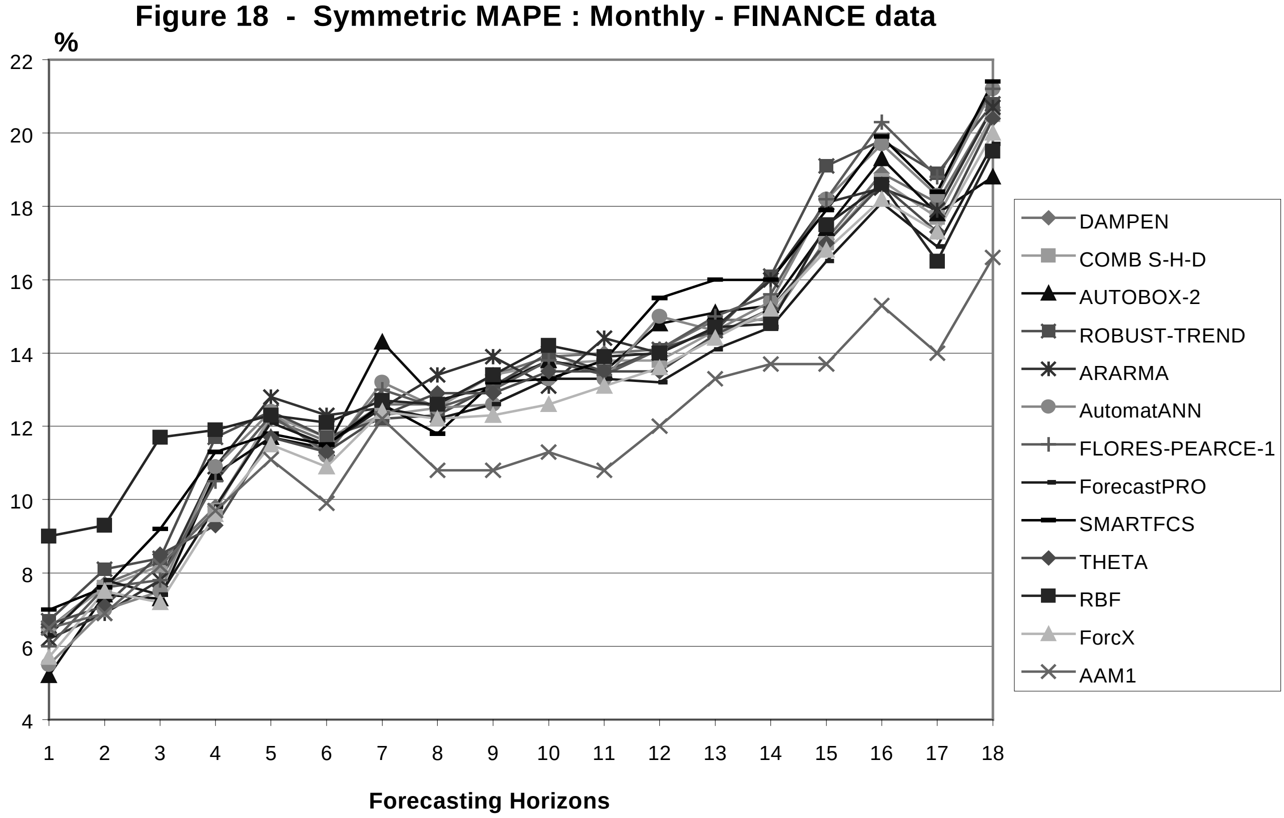


Figure 19 - Symmetric MAPE : Monthly - DEMOGRAPHIC data

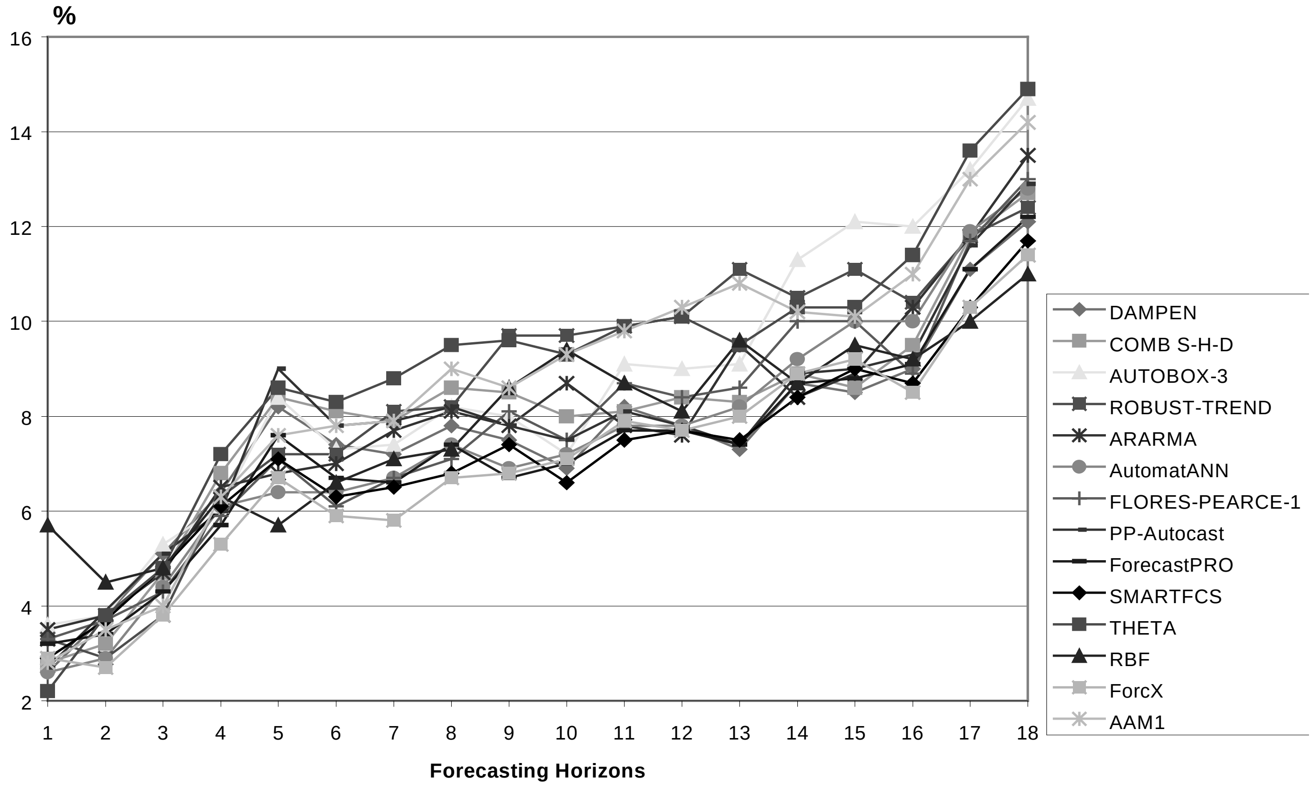


Figure 20 - Symmetric MAPE : Monthly - OTHER Data

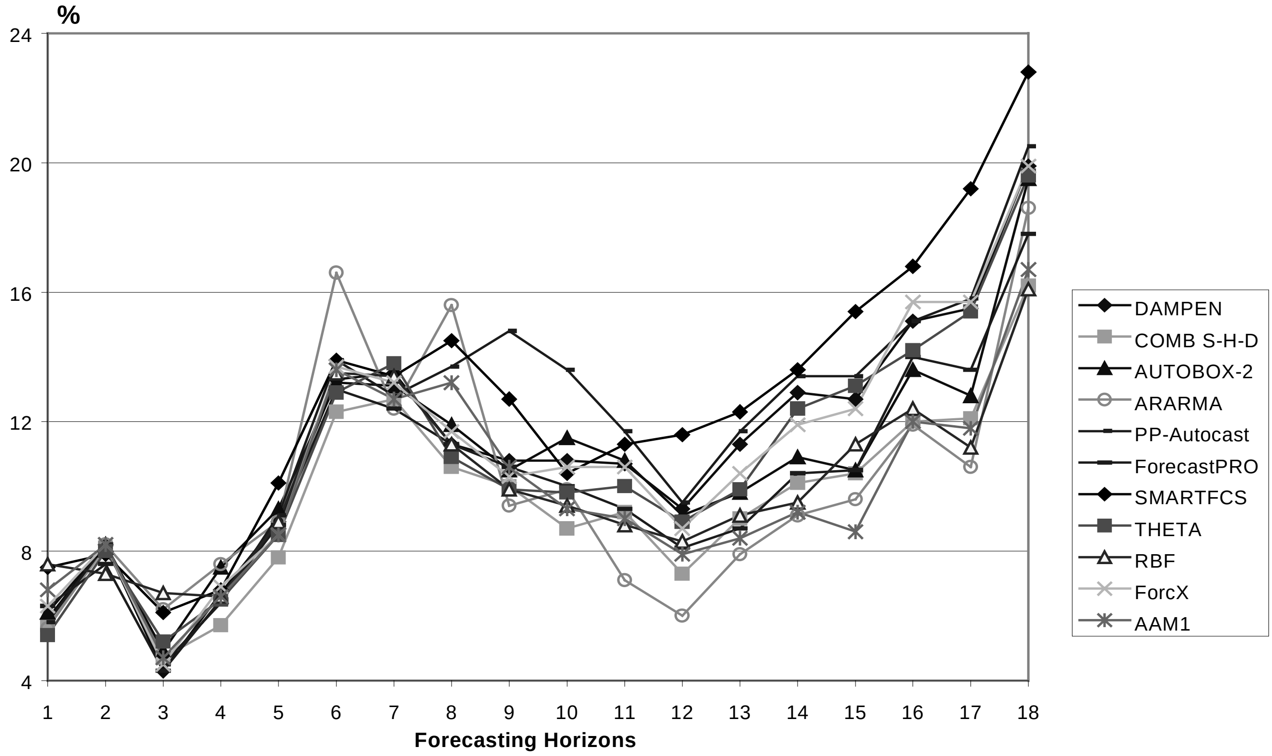


Table 1 **Methods which give the best results: Yearly Data**

Accuracy Measures	Micro (146)	Industry (102)	Macro (83)	Finance (58)	Demographic (245)	Total (645)
Symmetric MAPE	RobustTrend Flores-Pearc2 SMARTFcs Autobox2	THETA Comb-SHD Autobox2	RobustTrend ARARMA	Autobox2 SINGLE NAIVE2	ForcX RBF	RBF ForcX Autobox2 THETA RobustTrend
Average RANKING	RobustTrend THETA / Autobox2	THETA Comb-SHD / RobustTrend RBF	RobustTrend ARARMA RBF	SINGLE NAIVE2 / Autobox2 ForcX / ForecastPro	ForcX ForecastPro / PP-autocast	RBF / ForcX THETA / RobustTrend /Autobox2
Median APE	RobustTrend SMARTFcs	RobustTrend	RobustTrend ForecastPro	SINGLE NAIVE2 Autobox2	ForcX ForecastPro RBF THETA/ Autobox2	RBF FloresPearc1 PPautocast DAMPEN
Median RAE	RobustTrend SmartFcs / THETA / Autobox2	RobustTrend THETAsm THETA	RobustTrend ARARMA RBF		RBF THETA	RBF / THETA / RobustTrend Comb-SHD
RMSE	SINGLE NAIVE2 / RBF AutomatNN	RobustTrend THETA Comb-SHD	ARARMA RobustTrend Autobox3 / HOLT	RBF NAIVE2 SINGLE	THETA Comb-SHD ForecastPro	RBF RobustTrend SINGLE

Table 2 **Methods which give the best results: Quarterly Data**

Accuracy Measures	Micro (204)	Industry (83)	Macro (336)	Finance (76)	Demographic (57)	Total (756)
Symmetric MAPE	THETA Comb-SHD ForcX	Comb-SHD RBF ForcX PP-autocast	THETA Comb-SHD	THETA PP-autocast ForecastPro	THETA / SMARTFcs DAMPEN	THETA Comb-SHD DAMPEN PPautocast
Average RANKING	THETA HOLT Comb-SHD	Comb-SHD PP-autocast ForcX	THETA Comb-SHD DAMPEN	THETA ARARMA Comb-SHD	THETA / DAMPEN ARARMA	THETA Comb-SHD
Median APE	ForcX Comb-SHD HOLT	ForcX Comb-SHD THETA RobustTrend PP-autocast	THETA RBF FloresPearc1	THETA WINTER SMARTFcs	ARARMA RobustTrend	RobustTrend THETA Comb-SHD ForcX / DAMPEN PPautocast
Median RAE	HOLT THETA Comb-SHD/ RobustTrend	Comb-SHD/ THETA / RobustTrend HOLT	THETA / Comb-SHD	THETA / WINTER	THETA ARARMA Comb-SHD	THETA Comb-SHD RobustTrend
RMSE	THETA ForcX Comb-SHD / PP-autocast	NAIVE2 / Comb-SHD SINGLE	THETA SINGLE	THETA PP-autocast ForecastPro	SMARTFcs THETA FloresPearc2 Comb-SHD	THETA Comb-SHD

Table 3 **Methods which give the best results: Monthly Data**

Accuracy Measures	**Micro** (474)	Industry (334)	**Macro** (312)	**Finance** (145)	**Demographic** (111)	**Other** (52)	**Total** (1428)
Symmetric MAPE	THETA ForecastPro	ForecastPro ForcX BJ-automat	ARARMA RBF	AAM1 AAM2	ForcX SMARTFcs SINGLE ForecastPro	Comb-SHD BJ-automat AAM1	THETA ForecastPro
Average RANKING	THETA ForecastPro	ForecastPro ForcX THETA BJ-automat Comb-SHD	RobustTrend HOLT WINTER ARARMA / AAM1	AAM1 AAM2	RobustTrend	THETA AAM1 / AAM2 ARARMA / Comb-SHD	THETA ForecastPro Comb-SHD
Median APE	THETA ForecastPro	ForecastPro BJ-automat ForcX THETA	RobustTrend HOLT AAM1	AAM1 / AAM2 Autobox3 Autobox1	RobustTrend ARARMA / RBF	ARARMA AAM2	ForecastPro THETA
Median RAE	THETA THETAsm ForecastPro/ AutomtANN		AAM1 / RobustTrend HOLT ARARMA	AAM1 / AAM2	RobustTrend ARARMA	ARARMA AAM2 AAM1 THETA	
RMSE	THETA ForecastPro Forcx	BJ-automat ForecastPro ForcX	THETA Comb-SHD ForecastPro / DAMPEN	AAM1 / AAM2 AutomANN ForcX	SMARTFcs ForcX / SINGLE	BJ-automat ForecastPro AAM1 / Autobox2	ForecastPro ForcX THETA

Table 4 **Methods which give the best results: Other Data**

Accuracy Measures	**Micro**	Industry	**Macro**	**Finance** (29)	**Demographic**	**Other** (141)	**Total** (174)
Symmetric MAPE						THETA Autobox2 Comb-SHD / RobustTrend ARARMA	ARARMA THETA / Autobox2
Average RANKING				PPautocast DAMPEN		ForcX / Autobox2 RobustTrend THETA	ForcX / Autobox2 THETA ForecastPro/ RobustTrend
Median APE				AutomANN		ForcX Autobox2	ForcX / Autobox2 THETA / ForecastPro/ RobustTrend
Median RAE							
RMSE				DAMPEN PPautocast		Comb-SHD THETA	ARARMA

Table 5 **Methods which give the best results: Symmetric MAPE**

Time interval between Obs.	TYPES OF TIME SERIES DATA						
	Micro (828)	**Industry** (519)	**Macro** (731)	**Finance** (308)	**Demographic** (413)	**Other** (204)	**TOTAL** (3003)
Yearly (645)	RobustTrend FloresPearc2 SMARTFcs Autobox2	THETA Comb-SHD Autobox2	RobustTrend ARARMA	Autobox2 SINGLE NAIVE2	ForcX RBF		RBF ForcX Autobox2 THETA RobustTrend
Quarterly (756)	THETA Comb-SHD ForcX	Comb-SHD RBF ForcX PP-autocast	THETA Comb-SHD	THETA PPautocast ForecastPro	THETA **/** SMARTFcs DAMPEN		THETA Comb-SHD DAMPEN PPautocast
Monthly (1428)	THETA ForecastPro	ForecastPro ForcX	ARARMA RBF	AAM1 / AAM2	ForcX SMARTFcs SINGLE ForecastPro	Comb-SHD BJ-automat AAM1	THETA ForecastPro
Other (174)				DAMPEN / PPautocast AutomaANN ForecastPro		THETA Autobox2 RobustTrend Comb-SHD	ARARMA THETA / Autobox2
TOTAL (3003)	THETA ForecastPro	ForecastPro / ForcX THETA	RBF / ARARMA THETA / RobustTrend	AAM1 AAM2	ForcX		THETA ForecastPro

Table 6 **Methods which give the best results: Average RANKING**

Time interval Between Obs.	**TYPES OF TIME SERIES DATA**						
	Micro (828)	**Industry** (519)	**Macro** (731)	**Finance** (308)	**Demographic** (413)	**Other** (204)	**TOTAL** (3003)
Yearly (645)	RobustTrend Autobox2 THETA	THETA RobustTrend Comb-SHD RBF	RobustTrend ARARMA	SINGLE NAIVE2 / Autobox2 ForecastPro/ ForcX	ForcX PPautocast ForecastPro		RBF / ForcX THETA/ RobustTrend Autobox2
Quarterly (756)	THETA HOLT Comb-SHD	Comb-SHD PPautocast ForcX	THETA Comb-SHD DAMPEN	THETA ARARMA Comb-SHD	THETA / DAMPEN ARARMA		THETA Comb-SHD
Monthly (1428)	THETA ForecastPro	ForecastPro ForcX THETA Comb-SHD	RobustTrend HOLT WINTER ARARMA AAM1	AAM1 / AAM2	RobustTrend	THETA CombSHD ARARMA AAM1 / AAM2	THETA ForecastPro Comb-SHD
Other (174)				PPautocast DAMPEN		ForcX **/** Autobox2 RobustTrend THETA	Autobox2 ForcX THETA

Table 7 Methods which give the best results: Median APE

Time interval between Successive Obs.	TYPES OF TIME SERIES DATA						
	Micro (828)	**Industry** (519)	**Macro** (731)	**Finance** (308)	**Demographic** (413)	**Other** (204)	**TOTAL** (3003)
Yearly (645)	RobustTrend SMARTFcs	RobustTrend	RobustTrend ForecastPro	SINGLE NAIVE2 Autobox2	ForcX ForcastPRO RBF THETA Autobox2		RBF FloresPearc1 PP-autocast
Quarterly (756)	ForcX Comb-SHD HOLT	ForcX Comb-SHD THETA RobustTrend PPautocast	THETA RBF FloresPearc1	THETA WINTER SMARTFcs	ARARMA RobustTrend		RobustTrend THETA Comb-SHD ForcX
Monthly (1428)	THETA ForecastPro	ForecastPro BJ-automat ForcX THETA	RobustTrend HOLT AAM1	AAM1 / AAM2 Autobox3 Autobox1	RobustTrend ARARMA/ RBF	ARARMA AAM2	ForecastPro THETA HOLT Comb-SHD
Other (174)				AutomatANN		ForcX Autobox2	ForcX Autobox2 THETA ForecastPro

Table 8 Methods which give the best results: Median RAE

Time interval between Successive Obs.	TYPES OF TIME SERIES DATA						
	Micro (828)	**Industry** (519)	**Macro** (731)	**Finance** (308)	**Demographic** (413)	**Other** (204)	**TOTAL** (3003)
Yearly (645)	RobustTrend SmartFcs / THETA / Autobox2	RobustTrend THETAsm THETA	RobustTrend ARARMA RBF		RBF THETA		
Quarterly (756)	HOLT THETA Comb-SHD / RobustTrend	Comb-SHD/ THETA / RobustTrend HOLT	THETA / Comb-SHD	THETA / WINTER	THETA ARARMA Comb-SHD		
Monthly (1428)	THETA THEAsm ForecastPro / AutomatANN		AAM1 / RobustTrend HOLT ARARMA	AAM1 / AAM2	RobustTrend ARARMA	ARARMA AAM2 AAM1 THETA	
Other (174)							

Table 9 Methods which give the best results: RMSE

Time interval between Obs.	TYPES OF TIME SERIES DATA						
	Micro (828)	**Industry** (519)	**Macro** (731)	**Finance** (308)	**Demographic** (413)	**Other** (204)	**TOTAL** (3003)
Yearly	SINGLE NAIVE2 / RBF AutomatANN	RobustTrend THETA Comb-SHD	ARARMA RobustTrend Autobox3 / HOLT	RBF NAIVE2 SINGLE	THETA Comb-SHD ForecastPro		RBF RobustTrend SINGLE
Quarterly	THETA ForcX Comb-SHD PP-autocast	NAIVE2 / Comb-SHD SINGLE	THETA SINGLE	THETA PP-autocast ForecastPro	SMARTFcs THETA FloresPearce2 Comb-SHD		THETA Comb-SHD DAMPEN
Monthly	THETA ForecastPro ForcX	BJ-automat ForecastPro ForcX	THETA Comb-SHD ForecastPro DAMPEN	AAM1 / AAM2 AutomatANN ForcX	SmartFcs ForcX / SINGLE	BJ-automat ForecastPro AAM1 / Autobox2	ForecastPro ForcX THETA
Other				DAMPEN PP-autocast		Comb-SHD THETA	ARARMA THETAsm Autobox2

Table 10 Methods which give the best results: Symmetric MAPE - Monthly Data

Average Step horizons	TYPES OF TIME SERIES DATA						
	Micro (474)	**Industry** (334)	**Macro** (312)	**Finance** (145)	**Demographic** (111)	Other (52)	**TOTAL** (1428)
Short 1-3	SMARTFcs THETA ForecastPRO AutomaANN	ForecastPRO ForcX DAMPEN Comb-SHD THETA	Most of the methods	Autobox2 / AutomaANN ForcX	Most of the methods	Most of the methods	THETA ForecastPro SMARTFcs AutomANN ForcX
Medium 4-12	THETA ForecastPRO	ForecastPRO ForcX	Most of the methods	AAM1 / AAM2	Most of the methods	Comb-SHD BJ-automat	ForecastPro THETA ForcX
Long 13-18	THETA ForecastPRO	THETA ForcX / RBF ForecastPRO DAMPEN	RobustTrend RBF ARARMA AAM1	AAM1 / AAM2	SINGLE NAIVE2 / SMARTFcs ForcX / DAMPEN ForecastPro	AAM1 ARARMA RBF / Comb-SHD	THETA ForecastPro RBF
Overall 1-18	THETA	ForecastPRO ForcX	ARARMA RBF	AAM1 / AAM2	ForcX SMARTFcs SINGLE ForecastPro	Comb-SHD BJ-automat AAM1	THETA ForecastPro

Table 11 Methods which give the best results: Seasonal / Non-seasonal Data

	TYPES OF TIME SERIES DATA						
	Micro (828)	**Industry** (519)	**Macro** (731)	**Finance** (308)	**Demographic** (413)	**Other** (204)	**TOTAL** (3003)
Seasonal (862)	ForecastPRO THETA DAMPEN Comb-SHD SMARTFcs ForcX			AAM1 / AAM2 ForecastPRO ForcX			ForecastPRO THETA / Forcx / DAMPEN Comb-SHD
Non-Seasonal (2141)	THETA			AAM1 / AAM2			THETA ForecastPRO ForcX / Comb-SHD

APPENDIX:
Description of different new Methods

NAME OF THE METHOD: Robust Trend.

COMPETITOR: Nigel MEADE

DESCRIPTION:

This is a non-parametric version of Holt's linear model. The median based estimate of trend is designed to be uninfluenced by outliers. See Grambsch and Stahel (1990). The performance of the **Robust Trend** method agreed with that in Fildes and al (1996).

NAME OF THE METHOD: PPAutocast

COMPETITOR: Hans LEVENBACH

DESCRIPTION:

The method is the family of exponential smoothing methods associated with Ev.Gardner's work: Damped Trend Exponential Smoothing for Seasonal and Nonseasonal Time Series having periodicity from 1 (Annual), through 26 (Biweekly). We used 1, 4 and 12 as the seasonal periods for the M3 data. There are practical situations when 13 periods per year is relevant. As you know, single, Holt and Holt Winters are all special cases, since the damped trend models include no-trend, linear, and exponential. The particular model is data-driven, the algorithm searches automatically for the particular parameter set most relevant to the time series. Thus each of the M3 series has a unique set of parameters (which are part of the output file). The fitting criterion is MSE.

These models are integrated into PEER Planner for Windows which is a total forecasting system incorporating a GUI (Windows) interface, statistical forecasting engine, planned promotion models, MS Access relational database and review/override/reporting facilities.

In short, forecasting methodology is 100%, vanilla, implementation of Gardner's published methods. However, it is my experience with operational forecasting applications (product and inventory planning) is that the statistical methodology is about 20% of the achievable accuracy at best. The M3 is not very typical of the time series that need to be forecasted in operational situations (weekly and monthly data with promotion patterns, price effects, individual customer and market forces, etc. play a much bigger role. Hence the damped exponential smoothing methods (like the BJ models) serve only to capture the 'baseline' patterns in the demand history.

NAME OF THE METHOD: THETA-sm

COMPETITORS: P. MOURGOS and V. ASSIMAKOPOULOS

DESCRIPTION:

Theta-sm Model is a hybrid forecasting method, which is based on a successive filtering algorithm and a set of heuristic rules for both extrapolation and parameter calibration. The method focuses on the generation of a fitted line, which encompasses only the useful information for the information for the extrapolation.
An innovative feature of Theta-Model is that the fitting process relies on the identification of noisy and/or changing patterns in the original series.

There aren't any special conditions under which the model do better, but because of the above-mentioned innovative feature the model has a good performance in series characterized by changing patterns.

NAME OF THE METHOD : THETA

COMPETITORS: P. MOURGOS and V. ASSIMAKOPOULOS

DESCRIPTION:

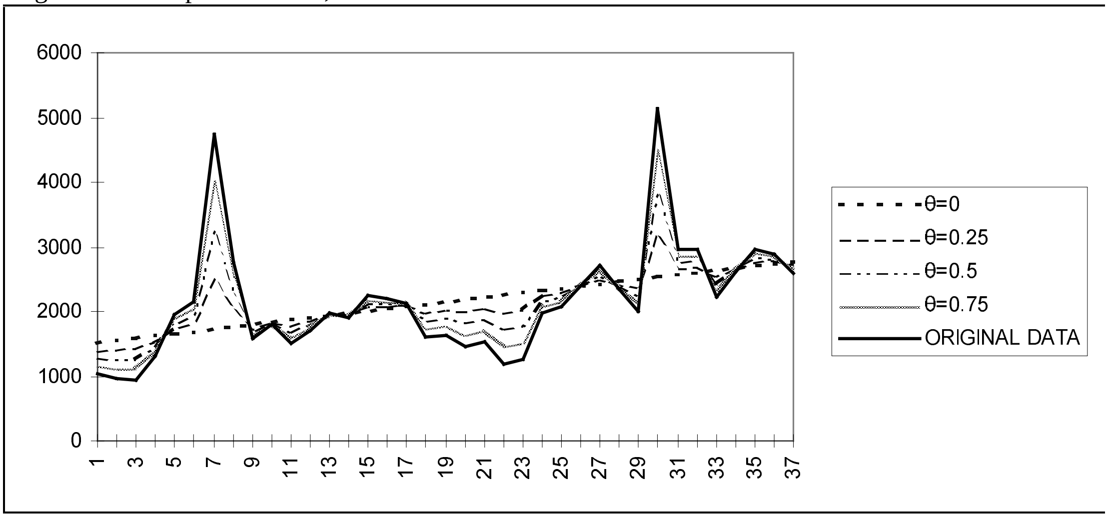
The model is based on the concept of modifying the local curvatures of the time series. The resulting series maintain the mean and the slope of the original data but not their curvatures.

This change is obtained from a coefficient, called θ-coefficient, which is applied directly to the second derivatives of the time series:

$$X''_{new} = \theta \cdot X''_{data}$$

If the local curvatures are gradually reduced then the time series is deflated as it is shown in Fig. 1. The smaller the value of the θ-coefficient, the larger the degree of deflation. In the extreme case where θ=0 the time series is transformed to a linear regression line. The progressive decrease of the fluctuations diminishes the absolute differences between successive momentary trends and is related, in qualitative terms, to the emerging of the data's long-term trends

Fig. 1. M3-Comp. Series 200, the θ-model deflation.



To the opposite direction, if the local curvatures are increased (θ>1), then the time series is dilated as it is shown in Fig. 2. The larger the degree of dilation, the larger the magnification of the short term behavior.

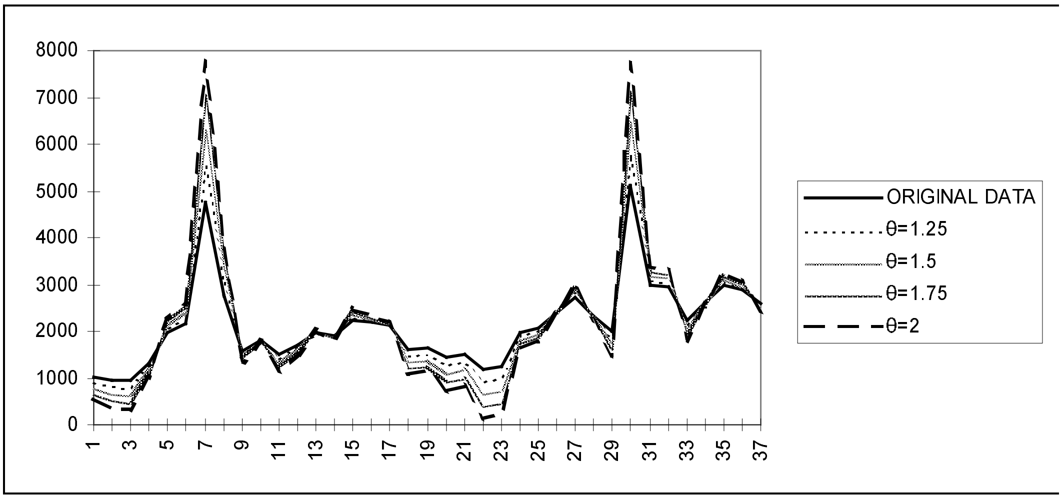


Fig. 2. M3-Comp. Series 200, the θ-model dilation.

The general formulation of the method becomes as follows:
The initial time series is disintegrated into two or more θ-lines. Each of the θ-lines is extrapolated separately and the forecasts are simply combined. Any forecasting method can be used for the extrapolation of a θ-line according to existing experience (Fildes and al., 1998). A different combination of θ-lines can be employed for each forecasting horizon.

Evaluation

The strong point of the method lies in the disintegration and transformation of the initial data. The two components include information, which is useful for the forecasting procedure but is lost or cannot completely be taken into account by the existing methods when they are directly applied to the initial data. Especially in the case of L(θ=0) this phenomenon is more comprehensible. The straight line includes information for the long-term trend of the time series which is "neglected" when a method tries to be adapted to the more recent trends. On the other hand, when the linear trend is used exclusively all the rest valuable information of the short term fluctuations is ignored.

The θ-model performance in the monthly time series of the M3 competition constitute a characteristic example. The monthly data of the competition were characterised, in general, by a relative large amount of volatility. This fact does not allow most methods to keep in memory the long-term trend and thus to take it into serious consideration in their forecasting function. In the case of θ-model the long-term trend is incorporated into the method as a major component through the L (θ=0) whereas its extrapolation is obvious by means of a simple continuation. At the same time, the existence of L(θ=2) operates as a counterbalance to the simplification of using a plain linear trend model. L(θ=2) increases the roughness of the monthly time series and augments the most recent trends. The effect of this augmentation is that the combined starting point reaches the "correct" level and since the extrapolation of L(θ=2) is horizontal the simple combination of both preserves a conservative but constant continuation of the long-term trend.

NAME OF THE METHOD: **AUTOBOX** Robust Arima
(Univariate Box-Jenkins with Intervention Detection)

COMPETITOR: David REILLY

DESCRIPTION:

In the absence of causal variables a time series can be described as an explicit function of it's own history (previous values) or dummy variables (0,1) which may take on the form of a pulse, level shift (step), seasonal pulse or time trend. Model identification optimizes the combination of these two possible components. The form of the non-stationarity is empirically identified, i.e. differencing or detrending. By combining stochastic, i.e. ARIMA structure with deterministic structure a more powerful estimation equation is possible. The unique contribution of AUTOBOX is to recognize that while a step is the finite sum of a pulse, a time trend is a finite sum of a step. This is incorporated into the modeling.

AUTOBOX was set up in a batch mode and conditions were set under which modeling was to be performed. Each time series was analyzed under four sets of conditions and the model selected was the one that minimized the error sums of squares.

The four conditions were:

1. Perform ARIMA modeling first then do INTERVENTION DETECTION, allowing local time trends to be identified.
2. Perform ARIMA modeling first then do INTERVENTION DETECTION, NOT allowing local time trends to be identified.
3. Perform ARIMA modeling second after INTERVENTION DETECTION, allowing local time trends to be identified.
4. Perform ARIMA modeling second after INTERVENTION DETECTION, NOT allowing local time trends to be identified.

The best of these four approaches was then declared the "winner" and its forecasts saved for submission to the M3-competition

CONDITIONS UNDER WHICH THE MODEL WILL DO WELL:

Non-causal forecasting does well when the omitted causal variables behave or arise consistently with their past.

COMMENTS:

The evaluation of a forecasting method, even with 3003 different time series still fails to provide generality due to the design of the competition. The single largest confusion in measuring and conducting forecasting competitions is the confusion between forecast errors from a single origin and forecast errors for different lead times. Single origin forecasts often generate a correlated set of forecasts due to the inherent bootstrapping procedures. That is to say the forecasts for one period out is correlated with the forecasts for two periods out, etc. thus the forecast errors are correlated.

To correctly measure forecast errors one has to compute k period projections from n origins. In this way one gets n independent measures of one period out errors, two period out errors , etc. . This requires an iterative process where the modeller is given a set and asked for a k period forecasts and is then given 1 new value and is asked to return another set of k period forecasts. In this way the effect of the origin or launch is designed out by virtue of the n replications. The developers of the M3 competition could have done this by scaling and coding these series thus masking the data and defeating any attempt to "cheat" .

A more important point is the flaw inherent in auto-projective models, i.e univariate models. The history of a series never causes or is responsible for the future. It is simply a surrogate for the omitted "cause" series. Box and Jenkins not only codified "rear-window driving" models (ARIMA) but developed a rigorous approach to causal modeling known as Transfer Functions. Transfer Functions are simply distributed lag models, which are optimally tuned to the data. By extracting the impacts or elasticities associated with casuals or exogenous series AND the history of the series one can project using the casuals rather than simply the rear-view mirror (ARIMA alone).

Until both of these issues are spoken to the question of which approach or model is optimal will remain unanswered.

NAME OF THE METHOD: **AAM** , automatic ARIMA modelling, with and without intervention analysis.

COMPETITORS: G. Mélard – J.M. Pasteels

DESCRIPTION:

We first restrict our analysis to the quarterly and monthly series (2184 series/3003). Yearly series have been discarding because some of them are too short to be modelled by the BJM. So, we choose not to treat them at all in order not to corrupt the sample series collected by the organizers. On the other hand, the series with unknown time interval between two successive observations (is it a day ?, an hour ?, a minute ?, a second ?) have been not considered because the BJM requires this piece of information in order to apply appropriate seasonal differences (7 or 5 for daily data, 24 for hourly data,…).

We have used an automatic ARIMA modelling system (see Mélard and Pasteels, 1995 and Pasteels, 1997). The system can be customized. There are about 20 commands lines concerning :

- the intervention analysis (maximum number, type of shocks to allow, treatment on the forecasting origin,…);
- the seasonal component (deterministic or not) ;
- the transformation criteria (setting the significance probability of the test) ;
- the differentiation (setting the significance probability of the test) ;
-…

We tried two modelling strategies (called respectively AAM and AAMi) described briefly below :

- AAM : classical Box-Jenkins methodology (no outlier treatment, no intervention analysis and stochastic seasonality) ;

- AAMi : same as AAM but with selective intervention analysis (for macro, industrial and demographic series, quarterly or monthly) ?

Expected performance, conditions under which it will do well :

We could at least expect the same performance than the ones observed for other competitions (see Fildes and Makridakis, 1993). For the M-Competition, AAM was performing well (for some horizons) for quarterly series and for series with a weak noise component. Whereas poor performance were obtained for noisy series (microeconomic) and more generally for horizons 1 and 2.

NAME OF THE METHOD: ARARMA

COMPETITOR: Nigel MEADE

DESCRIPTION:

The **ARARMA** methodology proposed by Parzen (1982) was applied with the benefit of human judgement (like the ARIMA models) in the M-Competition. The methodology used here was validated in Meade and Smith (1985) and automated for use in Fildes et al (1996). Apart from the transformation of the data to stationarity, Parzen preferring a long memory AR filter to the 'harsher' differencing used in ARIMA, a different approach to the identification of the ARMA model is used. Table 1 shows a comparison between Parzen's ARARMA forecasts and the procedure used here, the performance is broadly similar.

Table 1. Performance of ARARMA methods on 111 series sample of the M-Competition data

Horizon	M-Competition		ARARMA method used here	
	MAPE	MdAPE	MAPE	MdAPE
1	10.6	4.8	8.4	4.1
6	14.7	9.0	15.7	9.5
12	13.7	6.6	14.7	9.8
18	26.5	11.6	20.1	15.5

The following comments apply to both procedures. For seasonal series, the data was deseasonalised by routines provided by M. Hibon, the forecasts prepared and then reseasonalised. In order to distinguish between series that exhibit seasonality and those observations are merely monthly or quarterly the following procedure was adopted. The last six available observations were forecast out of sample under the assumptions that series was seasonal and that the series was non-seasonal. The assumption that provided the best Mean Absolute Percentage Error was used to provide the final forecast.

Fildes, R., M. Hibon, S. Makridakis and N. Meade, 1998, Generalising about Univariate Forecasting Methods: Further Empirical Evidence. *International Journal of Forecasting*, 14, 339-358

Grambsch, P., and W.A. Stahel, 1990, Forecasting Demand for Special Services, *International Journal of Forecasting*, 6, 53-64.

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.

NAME OF THE METHOD: ForecastPro

COMPETITORS : **R.GOODRICH, E. STEELWAGEN**

DESCRIPTION:

ForecastPro selects from among several methods: exponential smoothing and Box-Jenkins for mainstream data, Poisson and negative binomial models for low volume discrete data, Croston's method for intermittent data, and simple moving average for very short data sets. The selection process depends upon examination of the data and, in the case of exponential smoothing and Box-Jenkins, a rolling out-of-sample performance test.

CONDITIONS AND PERFORMANCE:

We try to cover all of the bulk of the data encountered in the business world by including several alternative models. Most of the M3 data appear to be fairly high volume, fairly long series, so almost all of the series will be forecasted via exponential smoothing or Box-Jenkins. We do best for fairly regular series but try to minimize losses (by switching to simpler models) when the series are highly irregular. Our methodology seems to perform best for monthly data.

NAME OF THE METHOD: SmartForecasts Automatic Forecasting System

COMPETITOR: Charles SMART

DESCRIPTION:

Smart Software's set of results submitted in the M3 Forecasting Competition was produced using the Automatic Forecasting expert system contained in **SmartForecasts** for Windows. The Automatic Forecasting system conducts a forecasting tournament among the following methods:

- simple moving average
- linear moving average
- single exponential smoothing
- double exponential smoothing
- Winters' additive and Winters' multiplicative exponential smoothing (if the data are seasonal).

For each method used in the tournament, the program uses a bisection search to converge automatically on those parameter values which minimize the mean absolute forecasting error for the method. The combination of method and parameter values that minimizes the mean absolute error wins the tournament and is selected as the optimal forecasting method.

An important strength of **SmartForecasts**' automatic forecasting process is that it computes out-of-sample forecast errors by sweeping repeatedly through the historical data, using some of the earlier data to develop its forecasting equations and testing the equations on ever more recent data (i.e., out-of-sample data). This procedure, known in the forecasting literature as *sliding simulation*, improves the reliability of the error estimates. All forecast errors (one step ahead, two steps ahead, etc.) are weighted equally in computing the mean absolute error. Calculation of the mean includes degree-of-freedom penalties initialized from the data.

Another strength of automatic forecasting is that the user can switch seamlessly from forecasting mode to judgmental adjustment mode. **SmartForecasts**' unique "Eyeball" feature lets you adjust statistical forecast results directly on-screen using a variety of "what if", goal-seeking and management override capabilities to reflect your knowledge and judgment. Full use is made of the interactive graphics available under Windows to make forecast adjustments and see both the forecast graph and numerical results change simultaneously. This combination of automatic statistical forecast generation plus optional judgmental adjustments can help to increase the accuracy and realism of your final forecast results.

NAME OF THE METHOD: R B F (Rule-based Forecasting)

COMPETITORS : **Monica Adya, Scott Armstrong, Fred Collopy, Miles Kennedy**

DESCRIPTION:

The forecasts for the M3-Competition were produced using Rule-Based Forecasting (RBF) as described in Collopy and Armstrong (1992) and Adya, Collopy and Kennedy (1997). Additional modifications to RBF were required to deal with the different types of series and the absence of domain knowledge. In this note we describe the differences between the original rule-base described in Collopy and Armstrong (1992) and the one used to produce the forecasts for the M3 competition.

The revisions to RBF involved corrections, simplification, automatic feature identification, and recalibration due to the absence of causal force information.

Conditions for RBF's Success

Series for which certainty is moderate to low, series in which the number of instabilities (changes in trend, step changes, outliers, etc.) are small, and series for which there is a moderate to high amount of causal knowledge. When these conditions are not met, we expect RBF to perform about as well as equal-weights combining.

For the annual data, about 49% of the series meet both the two relevant conditions where we expect RBF to perform well (since we did not code causal forces that condition is not relevant). Therefore, we expect that RBF will perform better than equal weights overall for the annual data.

This is our first real extension of RBF to quarterly, monthly, and other data. For these periods, equal-weights combining has not be as clearly the dominant option as it has been in studies of annual data. We anticipate that RBF will again do well relative to the component methods for the series where the above conditions are met. This is the case for about 49% of the quarterly and 69% of the monthly.

References

Adya, M., F. Collopy, and M. Kennedy, 1997(a), "Critical Issues in the Implementation of Rule-Based Forecasting: Evaluation, Validation, and Refinement", Working Paper, University of Maryland Baltimore County, Baltimore, MD.

Adya, M., F. Collopy, and M. Kennedy, 1997(b), "Heuristic Identification of Time Series Features: An Extension of Rule-based Forecasting", Working Paper, University of Maryland Baltimore County, Baltimore, MD.

Armstrong, J.S. and F. Collopy, 1993, "Causal Forces: Structuring Knowledge for Time Series Extrapolation", Journal of Forecasting, 12, 103-115.

Collopy, F. and J.S. Armstrong, 1992, "Rule-Based Forecasting: Development and Validation of an Expert Systems Approach to Combining Time Series Extrapolations", Management Science, 38, 10, 1394-1414.

NAME OF THE METHOD: FLORES-PEARCE

COMPETITORS: Benito FLORES and Steven PEARCE

DESCRIPTION:

The **Flores-Pearce** method utilizes an expert system that chooses among four forecasting methods based on the characteristics of the data. The system automatically determines whether or not the data have trend, and/or has periodicity. Then, it fits the most appropriate of : Simple Exponential Smoothing (SES), SES with Seasonality, Gardner's dampened trend or Gardner's dampened trend with seasonality.

Prior to characterizing the data and choosing a model, the series are examined for irrelevant early data and possible outliers. These, if present, are automatically removed. Because of the automatic nature of the system it should do better when there is a definite pattern on the data such as trend and periodicity or when irrelevant early data or outliers are present in the data.

The expert system is constructed using the C-Language integrated production system (CLIPS) developed by NASA. CLIPS uses a forward chaining method and is rule based. The rule set was developed from an examination of the rules available in the literature especially from the ones used by Collopy and Armstrong and based on the authors' own experiences. There are approximately 90 rules in the rule base.

As an added feature the expert system graphically presents the series and the forecasting model it has chosen to the user. The system allows the user to intervene in one of several ways. The user may select a different method from the choices, alter the forecast of the expert system method, or select another method and/or alter the forecast of this new method (or not).

It can be conjectured that the modified forecasts should do better if the data reveal pronounced trend that should not continue in the future or when the data shows a late change in the data pattern that has not yet being detected by the expert system rules. There is also the possibility that user modification can be beneficial when the expert system has identified an unusual periodicity, such as 11 for monthly data.

For the M3 competition the authors presented two sets of forecasts. One set generated by the **Flores-Pearce** expert system and another generated by the modification of the expert system forecasts by a user in the manner described above.

NAME OF THE METHOD: AutomatANN

COMPETITORS: Keith ORD, Sandy BALKIN

DESCRIPTION:

Artificial Neural Networks (**ANN**s) are an information paradigm inspired by the way the brain processes information. Using neural networks requires the investigator to make decisions concerning the architecture or structure used. **ANN**s are known to be universal function approximators and thus are capable of exploiting nonlinear relationships between variables.

This method, called ***Automated ANNs***, is an attempt to develop an automatic procedure for selecting the architecture of an artificial neural network for forecasting purposes.