The M-Competition Forecast Data

Matthias Klapper, Thomas Wenzel

Department of Statistics, University of Dortmund,

44221 Dortmund, Germany

klapper@amadeus.statistik.uni-dortmund.de

wenzel@amadeus.statistik.uni-dortmund.de

Abstract: The data used by Makridakis et al. (1982) is presented and explained. It contains forecasts that can be combined.

Keywords: M-competition, combination of forecasts.

Acknowledgement: Supported by Deutsche Forschungsgemeinschaft, Graduiertenkolleg „Angewandte Statistik“, Sonderforschungsbereich 475. M. Hibon kindly provided us with a copy of the data.

AMS 1991 Subject Classification: 62-00

1. Introduction

The data set was originally gathered for the famous M-competition presented in Makridakis et al. (1982). We were able to obtain the dataset from Michele Hibon, one of the co-authors. Chapter 2 describes the data sources and Chapter 3 the data structure. The computer code to read the data files is listed in Appendix A and a sample of the data in Appendix B. We did not use the Fortran code that came with the series due to the lack of a compiler. We used the SAS code to create a textfile that we were able to read into S-Plus for further analysis.

2. Data Sources

The selection of the 1001 time series is not random although a wide spectrum is covered including different sources, starting and ending dates typically located in the 1960s and 1970s. Among the time series are car production numbers, company sales, good pulp and aluminum production, paper prices, fruit exports, vehicle registrations, gross national product, price indices, change in stocks, national expenditures, population movement, and unemployment, just to name a few. The data is from various countries including Argentina, Austria, Belgium, Brazil, Denmark, Finland, France, Germany, Greece, India, Italy, Japan, Mexico, The Netherlands, Nigeria, Portugal, Switzerland, United Kingdom, and the United States.

Table 1 shows the data being subdivided into yearly, quarterly, and monthly data and also subdivided into categories. More than 60% of the time series are monthly data. The distribution among the different categories is mostly even though there are a little less Total Firm series.

Table 1: Types of Time-Series Data and Number of Series in Each Category

	Micro-Data				Macro-Data			
	Total Firm	Major divisions	Below major divisions	Industry	GNP	Below GNP	Demo-graphic	Total
Code	F	M	B	I	G	C	D	
Yearly	16	29	12	35	30	29	30	181
Quarterly	5	21	16	18	45	59	39	203
Monthly	10	89	104	183	64	92	75	617
Subtotal	31	139	132	236	139	180	144	1001
TOTAL	302			236	319		144	1001

3. Data Structure

For each time series there is one series of original data and 12 forecasts using different forecast methods, called aep, arrexp, bayes, brown, combinb, holt, movavg, naive, quadr, regres, single, and winter. Appendix B shows one quarterly, one monthly, and one yearly time series and the corresponding forecast series using method „holt“.

Preceding every time series is a data line giving technical information of the series. The first part is a three letter description of the time serie. The first variable is the name of the series, the first two letters standing for the frequency, e.g. YA for yearly, QR or QN for quarterly, and MR or MN for monthly data. The third letter corresponds with the columns in Table 1, e.g. F for „Firm“, M for „Major Divisions“, B for „Below Major Divisions“, I for „Industry“, G for „GNP“, C for „Below GNP“, and D for „Demographic“. The following variable is a running index. The next variable, i.e. 22 in the yearly series in Appendix B, is the number of observations which Makridakis et al. (1982) used in their well known M-competition as the past to generate the forecasts. After that the length of seasonality is explained, e.g. 1 for yearly, 4 for quarterly and 12 for monthly, followed by the number of future values forecasted. There are 6 forecasts for yearly time series, 8 for quarterly data, and 18 for monthly series. The last variable consists of two digits, the first classifying the series into yearly (1), quarterly (2 or 3) and monthly (4 or 5) data, the second digit being the column number (1 to 7) of Table 1. Below this explanatory row are the observations, first the past values and then the forecasted values followed by 12 seasonality indices. If this index is equal to one there is no seasonality and if it is equal to zero it is irrelevant. The last three lines give a brief description of the time series, e.g. starting and ending date, data source, etc.

The structure of the forecast files is much less complicated. The first row is identical with the first row in the original data file. This line is followed by the forecasts from a specific forecasting technique. As mentioned above there are 6 forecasts for yearly data, 8 for quarterly data and 18 for monthly data.

For combinations of forecasts we require several forecasts and a realization for the same event. Since we do only have forecasts for future values and no fitted values for the past, and we do not want to spend time on detailed time series analysis to recalculate these fits, we only use the monthly series that have 18 future values. For the analysis, we can now for example split those 18 data points into a „past“ of 10 points to calculate weights and a „future“ of 8 data points to look at the performance of different combination techniques.

Bibliography

- Makridakis, S., et al. (1982): The Accuracy of Extrapolation (Time-Series) Methods: Results of a Forecasting Competition, Journal of Forecasting Vol. 1, 111-153.

Appendix A: Computer code to read the data

A.1 Fortran

```
CHARACTER*5 TEXT1(15)
        CHARACTER*5 TEXT2(15)
        CHARACTER*5 TEXT3(15)
C
C       READ THE DATA
C
10      READ(20,100) FILE,N,L,NF,INDEX
100     FORMAT(A6,4I5)
        NT=N+NF
        NP1=N+1
        READ(20,200) (X(I),I=1,N)
        READ(20,200) (X(I),I=NP1,NT)
        READ(20,200) (D(I),I=1,12)
        READ(20,300) TEXT1,TEXT2,TEXT3
200     FORMAT(6E12.6)
300     FORMAT(15A5)
```

A.2 SAS

```
data test1;
  infile 'allser.dat';
  input name$ past type1 future type2;
  Do i=1 to past; input value e12. @; end;
  Do i=1 to future; input value1 e12. @; output; end;
  Do i=1 to 6; input; end; run;

data test2;
  infile 'aep.dat';
  input name$ past type1 future;
  Do i=1 to future; input value2 e12. @; output; end; run;

* Same procedure for arrexp.dat, bays.dat, brown.dat,
* combinb.dat, holt.dat, movavg.dat, naive.dat, quadr.dat,
* regres.dat, single.dat to create data sets test3-test12.

data test13;
  infile 'winter.dat';
  input name$ past type1 future;
  Do i=1 to future; input value13 e12. @; output; end; run;

data testxx;
  merge test1 test2 test3 test4 test5 test6 test7 test8
        test9 test10 test11 test12 test13; run;
  if substr(name,1,1)='M';
  file 'monforc.dat';
  put value1-value13; run;
```

Appendix B: The Data

B.1 Original Time Series

```
YAM22     22     1     6    12
0.542000E+02 0.651000E+02 0.970000E+02 0.915000E+02 0.867000E+02 0.113300E+03
0.111800E+03 0.111700E+03 0.103800E+03 0.791000E+02 0.785000E+02 0.963000E+02
0.932000E+02 0.811000E+02 0.842000E+02 0.874000E+02 0.883000E+02 0.793000E+02
0.736000E+02 0.839000E+02 0.100000E+03 0.102800E+03
0.935998E+02 0.905998E+02 0.105400E+03 0.118700E+03 0.112800E+03 0.122700E+03
0.100000E+01 0.000000E+00 0.000000E+00 0.000000E+00 0.000000E+00 0.000000E+00
0.000000E+00 0.000000E+00 0.000000E+00 0.000000E+00 0.000000E+00 0.000000E+00
OTHER     YEARLY    1977      MICRO2
STATISTICAL BUNDESAMT
INVENTORY PAPER WOOD PRICE IN WEST GERMANY
```

```
QNG19     60     1     8    35
0.617600E+04 0.597200E+04 0.609200E+04 0.611600E+04 0.617200E+04 0.615200E+04
0.642400E+04 0.645600E+04 0.643600E+04 0.647200E+04 0.649200E+04 0.637200E+04
0.647600E+04 0.674000E+04 0.679600E+04 0.692400E+04 0.678000E+04 0.676800E+04
0.674800E+04 0.688400E+04 0.686800E+04 0.671200E+04 0.716000E+04 0.710400E+04
0.760800E+04 0.776900E+04 0.802800E+04 0.808800E+04 0.833200E+04 0.838800E+04
0.826400E+04 0.833600E+04 0.848000E+04 0.835600E+04 0.840800E+04 0.850400E+04
0.872400E+04 0.860800E+04 0.863600E+04 0.903200E+04 0.921600E+04 0.928000E+04
0.937600E+04 0.957200E+04 0.985200E+04 0.101040E+05 0.103360E+05 0.106920E+05
0.106080E+05 0.110200E+05 0.109280E+05 0.109600E+05 0.112480E+05 0.113720E+05
0.116280E+05 0.119200E+05 0.114640E+05 0.116880E+05 0.116960E+05 0.120400E+05
0.122880E+05 0.125640E+05 0.125520E+05 0.128880E+05 0.128920E+05 0.131560E+05
0.136960E+05 0.138560E+05
0.100000E+01 0.100000E+01 0.100000E+01 0.100000E+01 0.000000E+00 0.000000E+00
0.000000E+00 0.000000E+00 0.000000E+00 0.000000E+00 0.000000E+00 0.000000E+00
OTHER     QUARTERLY 1971/4    MACRO1
OECD/OCDE 1960 -1975
GOVERNMENT EXPENDITURE CANADA
```

```
MRF1      42     1    18    41
0.697458E+06 0.118765E+07 0.106969E+07 0.107843E+07 0.105991E+07 0.105802E+07
0.512728E+06 0.884901E+06 0.905529E+06 0.617722E+06 0.410292E+06 0.854875E+06
0.652371E+06 0.937352E+06 0.151348E+07 0.113808E+07 0.687150E+06 0.136612E+07
0.108318E+07 0.137533E+07 0.456139E+06 0.124369E+07 0.922653E+06 0.848388E+06
0.861378E+06 0.145923E+07 0.178253E+07 0.133167E+07 0.167224E+07 0.126613E+07
0.129027E+07 0.192650E+07 0.843637E+06 0.133052E+07 0.139426E+07 0.950071E+06
0.675046E+06 0.219687E+07 0.133066E+07 0.894202E+06 0.197254E+07 0.165659E+07
0.117918E+07 0.193945E+07 0.630177E+06 0.913313E+06 0.161799E+07 0.622073E+06
0.915206E+06 0.128297E+07 0.126791E+07 0.120672E+07 0.140686E+07 0.688207E+06
0.516501E+06 0.137801E+07 0.840271E+06 0.437741E+06 0.556951E+06 0.700000E+06
0.100000E+01 0.100000E+01 0.100000E+01 0.100000E+01 0.100000E+01 0.100000E+01
0.100000E+01 0.100000E+01 0.100000E+01 0.100000E+01 0.100000E+01 0.100000E+01
OTHER     MONTHLY   1975/DEC  MICRO1
PRIVATE
COMPANY DATA USA
```

B.2 Forecast Series “holt”

```
YAM22     22     1     6
0.106213E+030.109625E+030.113036E+030.116448E+030.119860E+030.123272E+03
```

```
QNG19     60     1     8
0.121212E+05 0.122469E+05 0.123726E+05 0.124982E+05 0.126239E+05 0.127496E+05
0.128753E+05 0.130010E+05
```

```
MRF1      42     1    18
0.151503E+07 0.153391E+07 0.155279E+07 0.157167E+07 0.159055E+07 0.160943E+07
0.162831E+07 0.164719E+07 0.166607E+07 0.168495E+07 0.170383E+07 0.172270E+07
0.174158E+07 0.176046E+07 0.177934E+07 0.179822E+07 0.181710E+07 0.183598E+07
```