Reproduction of Forecasting Methods in the M Competitions

John Boylan, Zied Babai,
Mona Mohammadipour and Aris Syntetos

Lancaster University, Kedge Business School,
Middlesex University, Cardiff University

International Symposium on Forecasting, Santander, 22 June 2016

Agenda

- Motivation
- Scope of this Project
- Accuracy Measures (Naive 1)
- Deseasonalisation (Naïve 2)
- Method Specifications
  - Simple Moving Averages
  - Single Exponential Smoothing
- Discussion and Further Work

LCF

Motivation

Instead of imagining that our main task is to instruct a computer what to do, let us concentrate rather on explaining to *human beings* what we want a computer to do.

(Knuth, 1984)

Reproduction of forecasting methods is not possible unless a method is *fully* explained.
This includes the simplest forecasting methods.

LCF

Definition of Reproducibility

If results are reproducible, then independent researchers are able to obtain the same numerical results by repeating the original study using the same methods on the same data.

(Boylan et al, IJF, 2015)

Comment

It should be possible for independent researchers to reproduce or replicate *without* any additional information from the author(s) of the original study.

LCF

Objectives of this Study

M-Competitions

- Attempt to reproduce M-competition results
- Uncover any hidden assumptions
- Specify methods so fully that all simpler methods are reproducible
- Make available as a resource to researchers

More generally

- Identify areas to focus on in making forecasting results reproducible.
- Stimulate debate about journal policy in this area.

LCF

Clarification of Terminology

	Same methods	Different methods
Same data	**Reproduction** *Numerical Results*	**Re-Analysis**
Different data	**Replication** *Qualitative Conclusions*	

Why Reproduction?

Why should we reproduce?

- Increases reliability of numerical results
- Increases confidence in qualitative conclusions
- Reproduction is a necessary condition for replication

Why don't we reproduce?

- Data not available
- Methods not sufficiently clearly specified
- Considered too dull!

LCF

M-Competitions

M1 and M3 competitions

- Landmark forecasting competitions.
- Established agenda for empirical research.
- "Can result in replicability and objectivity".
  (Fildes and Makridakis, 1995)

Research Question: Are the studies reproducible?

- Data has always been available.
- Are methods sufficiently well specified?
- Surely, not too dull to reproduce!

LCF

Scope of Simple Methods to be Reproduced

Name of Method	M1	M3
Naive 1	✓	✓
Naïve 2	✓	✓
Simple Moving Averages	✓	
Single Exponential Smoothing	✓	✓
Adaptive Response Rate ES	✓	
Holt's Method	✓	✓
Brown's Linear Method	✓	
Linear Regression Trend	✓	
Damped Trend		✓
Holt-Winters (Multiplicative)	✓	✓
Combining Single-Holt-Damped		✓
Theta Method		✓

LCF

Sources for Methods and Measures

Methods

- M1 (JF, 1982) and M3 (IJF, 2000) papers
- Sources cited by M1 or M3 papers
  - Makridakis & Wheelwright (Book, 1978)
  - Gardner & McKenzie (MS, 1985)
- Simmons (IJF, 1986) – Naïve 2
- Ledolter & Abraham (1984) - Backcasting

Accuracy Measures

- M1, M3 papers, and Armstrong & Collopy (IJF, 1992) (MSE, MAPE, MedAPE, sMAPE)

LCF

Methodology

Two independent teams

- One team uses MATLAB, the other Visual Basic
- Each works using same sources of information
- Each documents any additional assumptions

Chairperson

- Exchange results via ‘chairperson’ (JB)
- Iterate process until agreement on results

Similar process as in Miller & Williams reproduction study (IJF, 2015)

LCF

Accuracy Measures (Naïve 1)

Median of all APEs or median of median (per series) APEs?

- Agreement with M1 for in-sample MedAPE if use "overall median" rather than "median of medians".

- Two methods give quite different results in-sample:
  MedAPE = 5.9% (Overall Median) v
  MedAPE = 5.4% (Median of Medians)

- Out-of-sample Overall Median (New) very close to M1 results but not exact. Occasional results do not agree to the first decimal place.

LCF

Naïve 2 Method (Deseasonalisation)

What is Naïve 2?

"… as Naïve 1 but the data are deseasonalized and then seasonalized. The seasonal indices for deseasonalizing and seasonalizing the data were done by the decomposition method of the ratio-to-moving averages. The specifics of this method can be seen in Makridakis and Wheelwright (1978) pp 94-100."

(Makridakis et al, 1982)

LCF

Early Attempt to Reproduce M1 Results for Naïve 2

Simmons (1986) found that:

- The seasonal indices used in Naïve 2 could not be reproduced using the procedure defined in the paper.
- The median was not calculated as one may have expected.

Simmons clarified, through written correspondence with Spyros Makridakis:

- A pre-processing procedure is executed before employing the decomposition method.
- Avoids losing second half of the last seasonal cycle.

LCF

Pre-Processing Procedure (Simmons, 1986)

Let X_t $(t = 1, \ldots, N)$ be defined as the original time series, L be defined as the number of seasons, and K be defined as the number of whole years in the data. Thus, K is the largest whole number less that or equal to N/L. Then Z_t $\{t = 1, \ldots, M$; where $M = (K + 1)L\}$, the actual series used by the centered ratio-to-moving average decomposition method, is defined in the following steps:

(1) $Z_t = (X_{N-3} + X_{N-2} + X_{N-1} + X_N)/4$, for $t = M - L/2 + 1, \ldots, M$,

(2) $Z_t = X_{t+N-M+L/2}$, for $t = L/2 + 1, \ldots, M - L/2$,

(3) $Z_t = (X_1 + X_2 + X_3 + X_4)/4$, for $t = 1, \ldots, L/2$ with the following exceptions:

$Z_t = X_{t-0}$ if $N - KL \geq L/2 - 0$, for $t = 1, \ldots, L/2$,

$Z_t = X_{t-1}$ if $N - KL = L/2 - 1$, for $t = 2, \ldots, L/2$,

$\vdots$

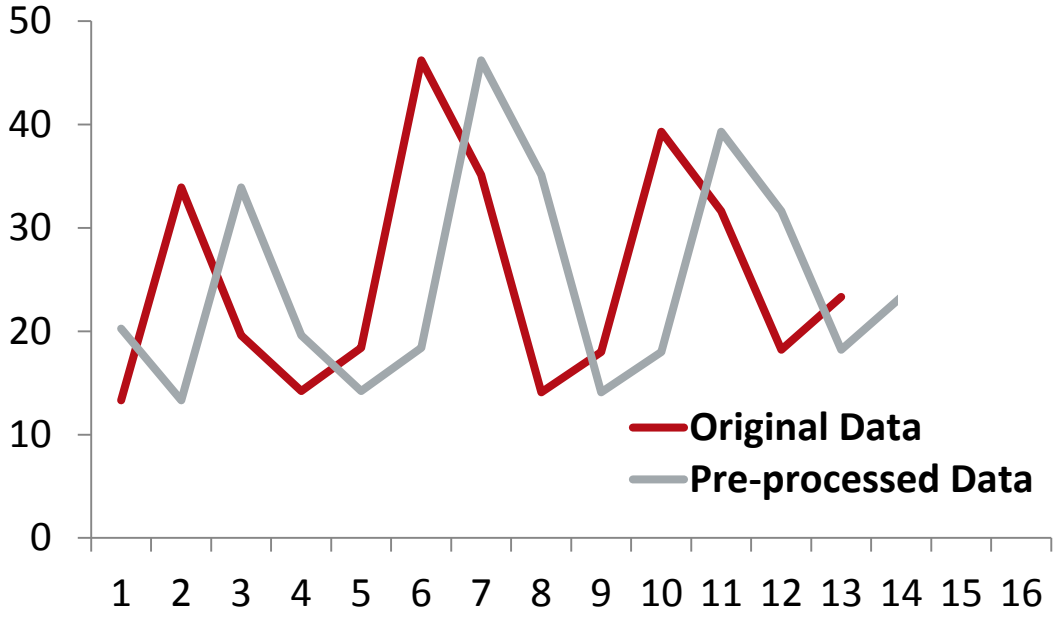
$Z_t = X_{t-(L/2-1)}$ if $N - KL = L/2 - (L/2 - 1)$, for $t = L/2, \ldots, L/2$.

Step (2) means that the new series is the same as original series shifted forward by *N-M+L*/2 periods.

LCF

Example of ‘Shifting’ of Series

Series QNM1 (13 in-sample observations)



Pre-processing shifts the data forward by one period .

Seasonal indices, based on the pre-processed data, should be applied to the previous period of the original data

LCF

Issues not Clarified by Simmons

- Length of data for pre-processing should be based on in-sample series.
- Inclusion of series with insufficient data to calculate medial (Winsorized) average – use straight average
- Application of seasonal indices to shifted data

- *Having clarified these aspects of the method, seasonal indices agreed with M1 indices.*

Specification of SMA in the M1 Paper

(2) Simple moving average.

$$\text{Model fitting: } \hat{X}_{t+1} = \frac{X_t + X_{t-1} + X_{t-2} + \cdots + X_{t-N+1}}{N}, \tag{6}$$

where N is chosen so as to minimize Σe_t^2, again summing over t from 1 to $n-m$

$$\text{Forecasts: } X_{n-m+k} = \frac{X_{n-m+k-1} + X_{n-m+k-2} + \cdots + X_{n-m+k-N}}{N}. \tag{7}$$

When the subscript of X on the right-hand side of (7) is larger than $n-m$, the corresponding forecasted value is substituted.

- This is not a 'flat forecast' throughout the horizon, based on the last *N* periods. The forecast for *t+2* includes the forecast for *t+1* in the Moving Average.
- Clarifications: minimise MSE (not SSE); N ≥ 2.

LCF

SMA Comparison (M1 Competition)

MAPE

Horizon	1	3	6	12	18
M1	13.0	17.6	25.0	17.8	31.3
M1Reproduced	13.0	18.0	25.0	18.5	30.3
M1Re-ananlysed	13.0	18.2	25.4	19.0	30.3

MedAPE

Horizon	1	3	6	12	18
M1	6.8	9.0	13.1	11.8	15.4
M1 Reproduced	7.6	9.3	12.8	12.1	15.3
M1 Re-analysed	7.6	9.5	13.1	12.6	16.2

- Reproduction: some differences remain.
- Re-analysis using ‘flat forecast’ shows further differences.

Specification of SES in the M1 Paper

(3) Single exponential smoothing.

$$\text{Model fitting: } \hat{X}_{t+1} = \alpha X_t + (1-\alpha)\hat{X}_t, \tag{8}$$

where α is chosen so as to minimise Σe_t^2, the mean square error where again summing is over t from 1 to $n-m$.

$$\text{Forecasts: } \hat{X}_{n-m+k} = \alpha X_{n-m} + (1-\alpha)\hat{X}_{n-m+k-1}. \tag{9}$$

- Again, this is not a 'flat forecast' throughout the horizon.
- According to Equation (9), the forecast for *t+2* includes the forecast for *t+1* in the update of SES.

LCF

Backcasting

- Backcasting as defined in Ledolter and Abraham (1984) is based on within-sample:
  - Same smoothing constant as for forecasting,
  - The initial value for backcasting is the last observation,
  - The last available smoothed value is used as the initial value for SES forecasting.

- This is the version of backcasting used to produce our results.

LCF

SES Comparison (M1 Competition)

MAPE

Horizon	1	3	6	12	18
M1	11.2	16.2	24.1	17.5	31.2
M1 Reproduced	11.2	16.5	24.8	16.6	30.9
M1 Re-analysed	11.2	16.3	24.2	17.5	30.1

MedAPE

Horizon	1	3	6	12	18
M1	6.0	8.6	12.5	11.3	14.9
M1 Reproduced	6.1	8.7	13.0	10.3	16.0
M1 Re-analysed	6.0	8.5	12.6	11.3	15.0

- Reproduction: some differences at longer horizons.
- Re-analysis using ‘flat forecast’ close to M1 results.

Discussion

- M Competitions represented a major advance in empirical forecasting and data transparency. Since then, method transparency has not developed as fast as it should.
- Re-examination of the M-competitions illustrates the need for greater clarity on:
  - Error measures across series
  - Seasonal adjustment procedures
  - Forecasting methods employed out-of-sample
- IMA Jnl of Management Mathematics introduced a ‘kite mark’ for reproducible papers. Time for others to follow?

LCF

Thank you for your attention!

Q&A?

John Boylan

Lancaster University

j.boylan@Lancaster.ac.uk

Lancaster University Management School | Lancaster Centre for Forecasting