ESANN'2008 proceedings, European Symposium on
Artificial Neural Networks - Advances in Computational Intelligence and Learning
Bruges (Belgium), 23-25 April 2008, d-side publi., ISBN 2-930307-08-0.

Do we need Experts for Time Series Forecasting?

Christiane Lemke and Bogdan Gabrys

Bournemouth University - School of Design, Engineering and Computing
Poole House, Talbot Campus, Poole, BH12 5BB - United Kingdom

Abstract. This study examines a selection of off-the-shelf forecasting and forecast combination algorithms with a focus on assessing their practical relevance by drawing conclusions for non-expert users. Some of the methods have only recently been introduced and have not been part in comparative empirical evaluations before. Considering the advances of forecasting techniques, this analysis addresses the question whether we need human expertise for forecasting or whether the investigated methods provide comparable performance.

1 Introduction

Time series forecasting has been an active area of research in the last decades. Numerous competitions and studies have been carried out, three extensive and recent ones are described in [1], [2] and [3]. Makridakis and Hibon present the results of a forecasting competition with 3003 time series in [1]. They found that complex models do not necessarily outperform simple ones, that forecasting performance depends on the accuracy measure and the forecasting horizon used and that combinations of forecasts outperform individual methods on average. Stock and Watson conducted an extensive study using 215 U.S. macroeconomic series in [2], comparing 49 linear and nonlinear forecasting methods with no clear-cut winner. Teräsvirta et al. re-examine the nonlinear methods used in [2] with different model parametrisation in [3], still with non-conclusive results.

In all of these studies, forecasting experts spent time and knowledge designing and tuning methods with different degrees of complexity only to come to the same conclusion: No single best method that works well on all time series can be identified. Consequently, in practical applications, one would need forecasting experts to investigate specific time series and suggest a forecasting model. However, the fact that experts with sufficient application-specific and forecasting expertise are mostly rare and expensive leads to the question of how much loss in forecast accuracy one might expect from failing to consult experts, but using off-the-shelf methods instead. This work provides an empirical comparison of one-step-ahead and multi-step-ahead forecasting methods that might be chosen by users who are not forecasting experts. Results are compared with outcomes of a recent forecasting competition ([4]).

The paper is organised as follows: Sections two and three introduce individual forecasting methods and forecast combinations investigated here. Section four describes methodology and results of the empirical study, section five concludes.

2 Forecasting

Eight time-series forecasting algorithms for one-step-ahead prediction have been investigated. Where suitable, the algorithms have been implemented with fixed parameters as well as using a grid search for parameter estimation.
Naïve forecast: The forecast is equal to the value of the last observation.
Moving average: The forecast is calculated as the arithmetic mean of the last k observations. The size of the time window is either set to 12 in order to account for yearly seasonality or determined by grid-searching k-values from 1 to 24 and choosing the k with the lowest in-sample mean squared error.
Single exponential smoothing: This method is given by the formula $\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t$, where y_t and $\hat{y}_t$ denote the observation and the forecast at time t, respectively. A grid search between 0 and 1 is employed to find the smoothing parameter α. Alternatively, it is set to the middle of the parameter range, 0.5.
Taylor's exponential smoothing: A modified dampened trend exponential smoothing was introduced in [5]. A growth rate and the level of the time series are estimated by exponential smoothing and then combined with a multiplicative approach. All parameters are determined by a grid search or again set to 0.5.
Polynomial regression: This algorithm fits polynomials of the orders one to six to the time series by regressing time series indices against time series values. The polynomial with the lowest error on a validation set is chosen. Alternatively, a polynomial of order four is used.
Theta-model: The Theta-model was introduced in [6] and applies a coefficient θ directly to the second order differences of a time series, resulting in different series with modified curvature. Adopting formulas from [7], two curves with $\theta = 0$ and $\theta = 2$ are used and their forecasts combined by a simple average.
ARIMA: Autoregressive integrated moving average models (ARIMA) according to [8] are models with an autoregressive and a moving average part, fitted to differenced data. The original series as well as its first and second order differences are submitted to the automatic ARMA selection process of a MATLAB toolbox ([9]), choosing the prediction with the lowest in-sample error. The same process is implemented with undifferenced series only.
Neural network: A feedforward neural network with one hidden layer containing 12 neurons, trained by a backpropagation algorithm with momentum has been implemented. Input variables are 12 lagged values of the time series. These characteristics have been selected based on findings of an extensive review of work using artificial neural networks for forecasting purposes by Zhang et al. in [10]. Ten neural networks have been trained and their predictions averaged to obtain the final forecasts.

The one-step-ahead forecasting methods have also been implemented in a multi-step version. The regression approach and the ARIMA model natively provide multi-step forecasting. All the other methods have been modified in two different ways: The *iterated* approach feeds the last prediction back to the model as an input for the next step. The *direct* approach trains n different models for a n-step prediction, directly on the multi-step problem.

ESANN'2008 proceedings, European Symposium on
Artificial Neural Networks - Advances in Computational Intelligence and Learning
Bruges (Belgium), 23-25 April 2008, d-side publi., ISBN 2-930307-08-0.

3 Forecast Combination

Several forecasts are often available for the same problem, which leads to the question whether or not some or all of the individual forecasts can be combined to obtain a superior forecast. An impressive amount of work has been done in this area, reviews and summaries can be found in [11], [12] and [13]. Five forecast combination methods are investigated here and will be described in the following paragraphs. For the one-step-ahead forecast, past performance is assessed using the mean squared error of a rolling window of the twelve latest observations and forecasts. For the multi-step forecast, past performance is calculated in a validation period prior to the testing set.
Simple average: This method averages all available forecasts.
Simple average with trimming: This algorithm averages individual forecasts as well, but only the best performing 80% of the models are taken into account.
Variance-based model: Weights for a linear combination of forecasts are determined using past forecasting performance ([14]).
Outperformance method: In [15], Bunn proposes to determine weights based on the number of times a method performed best in the past.
Variance-based pooling, two and three clusters: Past performance is used to group forecasts into two or three clusters by a k-means algorithm as suggested by Aiolfi and Timmermann in [16]. Forecasts of the historically better performing cluster are then averaged to obtain a final forecast.

4 Experiments

A data set consisting of 111 monthly empirical business time series with 52 to 126 observations has been obtained from the 2006/2007 Forecasting Competition for Neural Networks and Computational Intelligence [4]. The task was to predict 18 future values. Since the testing set will not be published before summer 2008, the last 18 observations of the provided time series are used for an out-of-sample error estimation. For the multi-step-ahead prediction, this reduction of the original length did not leave sufficient observations for the parameter estimation of the models in some cases, thus, only series with more than 60 values have been considered here.

The mean squared error (MSE) and the symmetric mean absolute percentage error (SMAPE) have been used for evaluating methods. The MSE is used in the majority of empirical studies, while the SMAPE was used for results of the forecasting competition. It is given by SMAPE $= \frac{1}{n}\sum_{t=1}^{n}\frac{|y_t-\hat{y}_t|}{(y_t+\hat{y}_t)/2} * 100$, where y_t is the observation and $\hat{y}_t$ the forecast. In the following tables for out-of-sample performances, some values are given with parentheses, indicating fixed parameter values for the regression, moving average and exponential smoothing methods, and lacking preprocessing for the ARIMA method. MSE values are given relative to the performance of the naïve forecast.

4.1 Results

Tables 1 and 2 give average out-of sample MSE and SMAPE values. For one-step-ahead forecasting, the ARIMA model gives the best performance, followed by Taylor's and single exponential smoothing. Except for the simple average and the outperformance model, combinations outperform individual models, with the three cluster variance-based pooling clearly being the best method.

Method	**SMAPE**	**MSE**
Naïve forecast	18.7 (-)	1.0 (-)
Moving average	15.9 (17.6)	0.78 (1.2)
Single exp. smoothing	15.9 (17.4)	0.77 (0.89)
Taylor's exp. smoothing	16.0 (17.4)	0.74 (0.88)
Regression	45.1 (217.7)	13.0 (95.4)
Theta model	16.1 (17.4)	0.79 (0.89)
ARIMA	15.6 (15.6)	0.68 (0.69)
Neural network	17.0 (-)	1.2 (-)
Simple average	15.6 (50.1)	0.87 (2.28)
Simple average with trimming	14.8 (15.7)	0.65 (0.70)
Variance-based weights	14.4 (15.4)	0.59 (0.64)
Outperformance model	17.2 (16.4)	13.9 (0.79)
Variance-based pooling, 2 clusters	14.2 (15.6)	0.61 (0.68)
Variance-based pooling, 3 clusters	13.8 (15.4)	0.59 (0.63)

Table 1: Average out-of-sample error values for one-step-ahead methods, numbers in parentheses present results for minimised tuning efforts, where applicable.

In general, direct approaches perform better than iterated approaches for multi-step-ahead forecasting when parameters are estimated by grid search, which can be explained by the accumulation of the errors in iterated methods. Neural networks, however, are an exception. Direct Taylor's exponential smoothing performs very well for both error measures, while an iterated neural network performs best in terms of the SMAPE. The ARIMA method does perform well again, but does not outperform the two approaches mentioned before. Simple average with and without trimming provides best results for forecast combinations.

In the result tables, the numbers in parentheses represent results from methods where parameter tuning efforts were minimised. It is interesting to look at the effect this approach has: While it always worsens performance for the one-step-ahead models, the impact is less clear for the multi-step-ahead models. For individual multi-step methods, some methods provide better performance compared to the same models with parameters estimated by grid search. For combinations, performance is even significantly improved for most of the cases, again with the variance-based pooling using three clusters being the best performing method. This result could be explained by parameter estimation errors and overfitting issues being a bigger problem for multi-step-ahead models. The

ESANN'2008 proceedings, European Symposium on
Artificial Neural Networks - Advances in Computational Intelligence and Learning
Bruges (Belgium), 23-25 April 2008, d-side publi., ISBN 2-930307-08-0.

combination of forecasts then produces a more stable and robust forecast.

Since the testing set of the NN3 forecasting competition will only be published in summer 2008, a comparison of the results presented here and the results of the competition is based on the out-of-sample error estimation for our experiments and the testing results from the competition published in [4]. This is disadvantageous for the results of this study, as fewer observations are available for model training. In the competition, the best method had an SMAPE of 15.18%. The best multi-step-ahead forecasting method, a directly trained neural network, obtains an SMAPE of 16.32%, which would result in rank four of 25. Variance-based pooling (three clusters) on methods with fixed parameters performed well for both error measures and produces an SMAPE of 17.08%, which would be ranked in the seventh place.

Method	**SMAPE**	**MSE**
Moving average (iterated)	20.26 (21.82)	1.83 (4.04)
Moving average (direct)	18.08 (18.09)	0.80 (0.80)
Single exp. smoothing (iterated)	22.54 (24.44)	0.86 (0.88)
Single exp. smoothing (direct)	20.07 (21.67)	0.82 (0.84)
Taylor's exp. smoothing (iterated)	31.15 (21.74)	1.70 (0.84)
Taylor's exp. smoothing (direct)	17.91 (23.93)	0.74 (0.86)
Theta model (iterated)	23.58 (23.63)	2.14 (1.96)
Theta model (direct)	22.05 (22.86)	2.09 (2.08)
Neural network (iterated)	16.32 (-)	0.90 (-)
Neural network (direct)	20.90 (-)	1.30 (-)
ARIMA	20.36 (19.21)	0.81 (0.81)
Regression	78.48 (524.1)	5.03 (1215)
Simple average	19.48 (88.33)	0.80 (9.45)
Simple average with trimming	19.52 (18.98)	0.76 (0.81)
Variance-based weights	24.52 (19.44)	1.11 (0.80)
Outperformance model	28.00 (25.69)	1.19 (1.60)
Variance-based pooling, 2 clusters	20.68 (17.98)	1.08 (0.78)
Variance-based pooling, 3 clusters	24.36 (17.08)	1.25 (0.75)

Table 2: Average out-of-sample error values for multi-step-ahead models. Numbers in parantheses present results with fixed parameters, where applicable.

5 Conclusions

This work empirically investigated simple approaches of forecasting time series and included both traditional methods as well as some recently introduced approaches. Generally, results of previous studies have been confirmed: Complex methods do not necessarily outperform simple ones and performance of methods differs according to error measures and forecasting horizons used. Regarding the performance of the methods, variance-based pooling with three clusters is the best choice for one-step-ahead forecasting. In the multi-step case, a neural network with an iterative approach works best for the SMAPE measure, while variance-based pooling with three clusters performs well on both error measures when used on individual methods with fixed parameters.

Furthermore, it can be concluded that easy-to-use off-the-shelf algorithms seem to provide a forecast accuracy that can very well compete with the performance of methods experts used in the NN3 forecasting competition. For the multi-step prediction problem, efforts to estimate parameters as opposed to just picking one and to perform preprocessing of the time series even appeared to be harmful in the experiments presented here.

References

[1] S. Makridakis and M. Hibon. The M3-Competition: Results, Conclusions and Implications. *International Journal of Forecasting*, 16(4):451–476, 2000.

[2] J.H. Stock and M.W. Watson. A Comparison of Linear and Nonlinear Univariate models for Forecasting Macroeconomic Time Series. In R.F. Engle and H. White, editors, *Cointegration, causality and forecasting. A festschrift in honour of Clive W.J. Granger*, pages 1–44. Oxford University Press, 1999.

[3] T. Teräsvirta, D. van Dijk, and M.C. Medeiros. *Linear Models, Smooth Transition Autoregressions, and Neural Networks for Forecasting Macroeconomic Time Series: A Reexamination*. Pontifícia Universidade Católica de Rio de Janeiro, 2004.

[4] NN3 Forecasting Competition [Online], 2006/2007. Available online: http://www.neural-forecasting-competition.com/ [13/06/2007].

[5] J. W. Taylor. Exponential Smoothing with a Damped Multiplicative Trend. *International Journal of Forecasting*, 19(4):715–725, October-December 2003.

[6] V. Assimakopoulos and K. Nikolopoulos. The Theta Model: A Decomposition Approach to Forecasting. *International Journal of Forecasting*, 16(4):521–30, 2000.

[7] R.J. Hyndman and B. Billah. Unmasking the Theta Method. *International Journal of Forecasting*, 19(2):287–290, 2003.

[8] G.E.P. Box and G.M. Jenkins. *Time Series Analysis*. Holden-Day San Francisco, 1970.

[9] Delft Center for Systems and Control - Software [Online], 2007. Available online: http://www.dcsc.tudelft.nl/Research/Software [13/06/2007].

[10] G. Zhang, B.E. Patuwo, and M.Y. Hu. Forecasting with Artificial Neural Networks: The State of the Art. *International Journal of Forecasting*, 14(1):35–62, 1998.

[11] R.T. Clemen. Combining Forecasts: A Review and Annotated Bibliography. *International Journal of Forecasting*, 5:559–583, 1989.

[12] C.W.J. Granger. Invited Review: Combining Forecasts - Twenty Years Later. *Journal of Forecasting*, 8:167–173, 1989.

[13] A. Timmermann. Forecast Combinations. In G. Elliott, C.W.J. Granger, and A. Timmermann, editors, *Handbook of Economic Forecasting*, pages 135–196. Elsevier, 2006.

[14] P. Newbold and C.W.J. Granger. Experience with Forecasting Univariate Time Series and the Combination of Forecasts. *Journal of the Royal Statistical Society. Series A (General)*, 137(2):131–165, 1974.

[15] D.W. Bunn. A Bayesian Approach to the Linear Combination of Forecasts. *Operational Research Quarterly*, 26(2):325–329, June 1975.

[16] M. Aiolfi and A. Timmermann. Persistence in Forecasting Performance and Conditional Combination Strategies. *Journal of Econometrics*, 127(1-2):31–53, 2006.