# Forecasting and Forecast Combination in Airline Revenue Management Applications

Christiane Lemke[1], Bogdan Gabrys[1]

[1] School of Design, Engineering & Computing, Bournemouth University, United Kingdom.
E-mail: {clemke, bgabrys}@bournemouth.ac.uk

**Abstract.** Predicting a variable for a future point in time helps planning for unknown future situations and is common practice in many areas such as economics, finance, manufacturing, weather and natural sciences. This paper investigates and compares approaches to forecasting and forecast combination that can be applied to service industry in general and to airline industry in particular. Furthermore, possibilities to include additionally available data like passenger-based information are discussed.

**Keywords.** Forecasting, Forecast Combination, Adaptive Forecasting, Airline Industry

## 1 Introduction

Modern revenue management systems significantly increase revenues of airline companies. Airline tickets are usually sold for several booking classes differing in price and booking conditions. Passengers buying higher class tickets are willing to pay a higher price and thus contribute more to airline revenues than low fare passengers, which is why airlines would like to give priority to them. However, those higher class bookings usually arrive quite shortly before departure, so it becomes necessary to forecast the demand for higher class tickets to be able to reserve an appropriate number of tickets.

Two variables are important for this task: the demand and the cancellation rate for airline tickets. A collaboration between Bournemouth University and Lufthansa Systems Berlin in a previous project showed promising results and significant improvements for forecasting the demand using novel forecast combination approaches ([1], [2], [3]). A new project whose first results are presented in this paper aims at investigating if similar improvements can be achieved for cancellation rates.

### 1.1 Airline Revenue Management

Revenue management tries to maximise profits by investigating and forecasting customer behaviour and drawing appropriate conclusions. In the airline industry, the central objective of revenue management is determining how many seats for each booking class should be sold prior to departure. The risk of rejecting a booking in a low class in order to wait for a higher class booking has to be judged. Forecasting of ticket demand with corresponding no-show and cancellation rates are a crucial component to this. McGill and van Ryzin give an extensive review of research in airline revenue management [4]. One of the fundamental developments described is abandoning traditional single-leg approaches only considering individual flights in favour of so called Origin and Destination (O&D) approaches, which work combinations of all connections, itineraries and booking classes.

Forecasting has a long track record in airline industry, mainly because forecasting future demand has a direct influence on the booking limits for the different fare classes [4]. Fully occupied flights are of both ecological and economical interest; but a number of seats usually stay unoccupied even on sold-out flights due to cancellations or so called no-show passengers. Forecasting in this context is a crucial help for a reasonable overbooking, finding a balance between the number of unoccupied seats and the number of denied boardings.

The state-of-the-art method for forecasting demand and cancellation rates uses a statistical model that takes the currently observed rate and the reference rates obtained from historically similar flights into account. Several approaches can be pursued when trying to improve a forecast; two have been identified in the scope of this work: employing and improving traditional time series methods and extracting information about customer behaviour.

### 1.2 Traditional time-series forecasting

Forecasting and forecast combination is a well-researched area ([5], [6], [7]). Time series forecasting looks at sequences of data points, trying to identify patterns and regularities in their behaviour that might also apply to future values. A large number of time series forecasting methods with different degrees of flexibility and complexity are available; consequently, there are many ways to generate forecasts and one might end up with more than one forecast for the same problem. This leads to the question, whether or not some or all of the individual forecasts can be combined to obtain a superior forecast. General forecasting and forecast combination methods are discussed in the sections two and three of this paper, section four gives an empirical evaluation of popular and easy-to-use approaches.

### 1.3 Passenger-based predicting

In many real world applications, information that goes beyond ordinary time series data is often available. In services industry, data is often collected on a customer basis and can be utlised with data mining methods that help modelling and understanding various groups of customer behaviour. For example, data mining is used for customer relationship management in retail industry ([8]), for credit card fraud detection ([9]) or for market basket analysis identifying associations between buying choices of customers ([10]).

In the last years, a few publications suggest that including information gained from so-called Passenger Name Records (PNR) in airline forecasting applications might be beneficial. In [11] and [12], Neuling et al. and Lawrence et al. look at forecasting no-show rates, i.e. the rate of passengers who book a ticket and fail to show up for a flight without cancelling it. Both publications find that making use of PNR data improves forecasting performance compared to pure time series approaches.

## 2 Forecasting

This section describes a selection of popular time series forecasting techniques inspired by a book of Makridakis et al. ([5]). It provides the background for the further sections of this paper. In the formulas used, the time series will have the past observations $\{y_1,..,y_t\}$ and the one-step-ahead forecast to calculate will be denoted by $\hat{y}_{t+1}$.

### 2.1 Averaging and smoothing

A simple approach to forecasting is taking the arithmetic mean of the $k$ most recent values of the time series as shown in formula (2-1). In that way, old and potentially inapplicable values can be discarded.

$$\hat{y}_{t+1} = \frac{1}{k} \sum_{i=t-k+1}^{t} y_i \qquad (2\text{-}1)$$

Exponential smoothing methods apply weights that decay exponentially with time and thus also rely on the assumption that more recent observations are likely to be more important for a forecast than those lying further in the past. Single exponential smoothing as the simplest representative of smoothing methods is calculated by the previous forecast adjusted by the error it produced, see formula (2-2). The parameter $\alpha$ controls the extent of the adjustment.

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t \qquad (2\text{-}2)$$

Many extensions and variations to the basic smoothing algorithm have been proposed, see for example [13].

### 2.2 Regression

Regression approaches express a forecast or dependent variable as a function of one or more independent or explanatory variables that relate to the outcome. Simple linear regression on a single variable $x$ can be expressed with formula (2-3), where $a$ is the intercept, $b$ the slope of the line and $\varepsilon$ the error, that originates from the deviation of the linear relationship from the actual observation.

$$y = a + bx + \varepsilon \qquad (2\text{-}3)$$

In the case of time series, $x$ denotes the time index. The parameters of the regression can be estimated using standard least squares approaches.

### 2.3 Decomposition and Theta-Model

Decomposition aims at isolating components of a time series, projecting them separately into the future and then recombining them to produce a final forecast. The components are traditionally

- a trend-cycle, denoting long-term changes,
- seasonality, reflecting shorter-term constant-length changes like months or holiday times and
- an irregular or random error component.

Recently, Assimakopoulos and Nikolpoulos proposed the Theta-model in [14]. It decomposes seasonally adjusted series into short and long term components by applying a coefficient $\theta$ to the second order differences of the time series as shown in formula (2-4), thus modifying the curvature of the time series.

$$y''_{new}(\theta) = \theta * y''_{original} \tag{2-4}$$

Theta values bigger than one dilate the series, amplifying its short term-behaviour, while theta values between zero and one have the opposite effect.

### 2.4 ARIMA

Autoregressive integrated moving average models (ARIMA) according to [15] are a powerful and complex tool of modelling and forecasting time series. They are described by the notation ARIMA (p,d,q) and consist of the following three parts:

1. AR(p) denotes an autoregressive part of order $p$, autoregression is a regression where the target variable depends on time-lagged values of itself.
2. I(d) defines the degree of differencing involved. Differencing is a method of removing non-stationarity by calculating the change between each observation.
3. MA(q) indicates the order of the moving average part of the model, which is given as a moving average of the error series. It can be described as a regression against the past error values of the series.

### 2.5 Nonlinear Forecasting

Regime switching models are a popular class of nonlinear forecasting methods, combining two or more sets of model coefficients in one system. Which set to apply for a forecasting situation is then determined by looking at the regime or state the system is likely to be in. A simple example for a model with two autoregressive regimes of order one and the parameters $\alpha_i, \beta_i, \chi$, the error component $\varepsilon_i$ and the observable variable $z$ that governs the change of regimes can be found in formula (2-5).

$$\begin{aligned} y_t &= \alpha_1 + \beta_1 y_{t-1} + \varepsilon_1, \quad z \le \chi \\ y_t &= \alpha_2 + \beta_2 y_{t-1} + \varepsilon_2, \quad z > \chi \end{aligned} \tag{2-5}$$

Smooth switching of regimes in so called smooth transition autoregressive (STAR) models addresses issues that arise from abruptly changing regimes in the simple model. A survey of recent developments in this area is given in [16].

Regime switching models are a model-driven approach. Artificial neural networks can be used for data-driven forecasting, with the advantage of not having to choose an appropriate model for each problem. For time series forecasting, time lagged observations and time indices can be used as input variables, obtaining the forecast as an output. Neural networks have been frequently and successfully used for forecasting purposes, a summary of work done in this area can be found in [17].

### 2.6 Discussion

Time series forecasting has been extensively researched in the last 40 years and a large number of empirical studies have been conducted to compare out-of-sample accuracy of various methods. Among the biggest forecasting competitions are the three so called M-competitions, consisting of the M-competition 1982, the M2-competition in 1993 and the M3-competition in 2000. All three of them confirmed the same general results, as summarised in [18].

1. Statistically complex models like ARIMA do not necessarily outperform less sophisticated approaches like exponential smoothing.
2. Forecasting performance depends on the accuracy measure used.
3. Forecasting performance depends on different time horizons.
4. Combinations of forecasts do on average outperform the individual methods.

Especially conclusion number one has been subject to fierce discussions, many of them disagreeing on the fundamental question of whether or not empirical evaluations are an appropriate measure for the performance of a model. In [18], Makridakis and Hibon strongly criticise the approach of building statistically complex models, disregarding all empirical evidence that simpler ones predict the future just as well or even better in real life situations. In [5] it is furthermore added, that the only advantage a sophisticated model has compared to a simple one is the ability to better fit historical data, which is no guarantee for a better out-of-sample performance. In [19], Sandy D. Balkin criticises the choice of the M3-competition data sets as originating mostly from financial and economic time series and probably not containing enough complex structures that could favour more sophisticated models. Keith Ord adds in the same publication, that ARIMA models need at least 50 observations to be efficiently estimated, which is a requirement that is often violated.

Another extensive empirical study has been carried out by Stock and Watson in [20], using 215 U.S. macroeconomic series, comparing 49 linear and nonlinear forecasting methods. Looking at the nonlinear methods, they found that neural networks performed well for one-period-ahead forecasting, while showing deteriorating performance with increasing forecasting horizons. A smooth transition autoregression model performed generally worse than the neural networks. No clear-cut winner could be identified comparing nonlinear methods to linear ones, as the forecasting accuracy differed significantly across forecast horizons and series. In [21] however, the results of this study are questioned by stating that nonlinear forecasting methods should only be considered at all if the data shows nonlinear characteristics. They furthermore criticise carelessness in the parameterisation of the nonlinear methods being examined in the study. A re-examination of the performances of linear and nonlinear approaches on time series that rejected the statistical test for linearity has consequently been carried out using seven macroeconomic time series. The results, albeit carefully considering suitable model parameterisation, do not overly favour nonlinear models. A STAR model and one of the two examined neural networks had a slightly better performance than linear models, the second neural network had not.

Summarising, the results of the enormous amount of empirical studies have been mixed and sometimes contradictory; no single best general method could be clearly identified. In ([19]), Keith Ord suggests a rough guideline which method to choose. Generally, a small number of observations, very erratic process behaviour and no or weak seasonal pattern for a given time series are strong indicators that simple methods should be used. As the number of observations grows and the series exhibit a stable stochastic and a strong seasonal pattern, statistical criteria and contextual information should be used to identify an appropriate, possibly sophisticated model.

## 3 Forecast Combination

Since the publication of the seminal paper on forecast combination by Bates and Granger in 1969 ([22]), research in this area has been active; recent reviews and summaries can be found in [6] and [7]. In general, four main reasons for the potential benefits of forecast combinations have been identified:

1. It is implausible to be able to correctly model a true data generation process in only one model. Single models are most likely to be simplifications of a much more complex reality, so different models might be complementary to each other and be able to approximate the true process better.
2. Even if a single best model is available, a lot of specialist knowledge is required in most cases to find the right functions and parameters. Forecast combinations help achieving good results without in-depth knowledge about the application and without time consuming, computationally complex fine-tuning of a single model.
3. It is not always feasible to take all the information an individual forecast is based on into account and create a superior model, because information may be private, unobserved or provided by a closed source.
4. Individual models may have different speeds to adapt to changes in the data generation process. Those changes are difficult to detect in real time, which is why a combination of forecasts with different abilities to adapt might perform well.

This section presents forecast combination methods and provides a summary and discussion of empirical evaluations to assess their quality.

### 3.1 Linear Forecast Combination

The linear combination of forecasts calculates a combined forecast $\hat{y}^c$ as the weighted sum of $m$ individual forecasts $\{\hat{y}_1,..,\hat{y}_m\}$ as shown in equation (3-1).

$$\hat{y}^c = \sum_{i=1}^{m} \omega_i \hat{y}_i \qquad (3\text{-}1)$$

Weights can be estimated in various ways. One easy and often remarkably robust example is the simple average combination with equal weights.

The optimal model proposed in [22] calculates weights using formula (3-2), where $\Sigma$ denotes the covariance matrix of the $m$ different forecast errors and $e$ the $n \times 1$ unit vector.

$$\omega = \frac{\Sigma^{-1} e}{e' \Sigma^{-1} e} \qquad (3\text{-}2)$$

A variance based approach first mentioned by Bates and Granger in [22] and further extended by Newbold and Granger in [23] uses the average of the sum of the past squared forecast errors ( $MSE$ ) over a certain period of time as shown in formula (3-3).

$$\omega_i = \frac{MSE_i^{-1}}{\sum_{j=1}^{m} MSE_j^{-1}} \qquad (3\text{-}3)$$

Granger and Ramanathan propose the regression method in [24] and treat individual forecasts as regressors in an ordinary least squares regression including a constant.

Another group of linear forecasting methods does not estimate a covariance matrix or rely on past error values. In a rank-based approach according to Bunn ([25]), each combination weight is expressed as the likelihood that the corresponding forecast is going to outperform the others, based on the number of times where it performed best in the past. Gupta and Wilton additionally consider relative performance of other models using a matrix with pair-wise odd ratios in [26]. Each element of the matrix represents the probability that the model of the corresponding line will outperform the model on the corresponding column.

### 3.2 Nonlinear Forecast Combination

Potentially nonlinear relationships among forecasts are not considered in linear forecast combination, providing the main argument for usage of nonlinear combination methods.

The most investigated nonlinear method for forecast combination are back-propagation feedforward neural networks, where individual forecasts are input data and the combined forecast is obtained as the output. This method was first mentioned by Shi and Liu in [27] and was also used in [28] and [29].

Fuzzy systems for forecast combination can be found following two different paradigms. First, fuzzy systems can be seen as a kind of regime model similar to the one described in section 2.5, where two or more different forecasting models can be active at one time ([30]). In an empirical evaluation in the same paper, the resulting fuzzy system almost always outperforms or draws level with the individual forecasts and linear forecast combination methods investigated. Two more publications emphasise a different aspect of fuzzy systems - the possibility of modelling linguistic and subjective knowledge ([31], [32]). Combining expert forecasts with traditional time series forecasts resulted in significant performance gains in both publications.

In [33], He and Xu present a self-organizing algorithm based on the Group Method of Data Handling (GMDH) method which was proposed by Ivakhnenko in the 1970s

([34]). Individual forecasts are taken as an input variable for the algorithm, different transfer functions, usually polynomials, then create intermediate model candidates for the first layer. Iteratively, the best models are selected with an external criterion and used as input variables for the next layer, producing more complex model candidates until the best model is found.

Several authors favour the approach of pooling forecasts before combining them. By grouping similar forecasts and subsequently combining the pooled forecasts, several issues like increased weight estimation errors because of a high number of forecasts to combine can be addressed. Research in this area recently started with clustering forecasts based on their recent past's error variance in [35] and continued with investigations by Riedel and Gabrys on how to extend and modify the clustering criteria in the context of a big pool of individual forecasts that have been diversified by different methods in [2] and [3]. The tree-like structures of these multi-level multi-step forecast combinations can be evolved with genetic programming, using the quality of the combined predictions on the validation data as the fitness function to optimize.

### 3.3 Adaptive forecast combination

A constantly changing environment is a typical characteristic for an area in which forecasts are applied. Assuming that no individual model can be a perfect model of the true data generation process and considering that each individual model has a different speed to adapt to changes, there is a reason to believe that forecast combination will perform well wherever adaptivity is needed.

Taking into account the fact that the performance of individual forecasts changes over time, time varying combination weights have been investigated. One of the initial papers on that matter [36] proposes modelling a bigger impact of more recent observations and letting the combination weights be a function of time. One plausible method in the context of regression and variance-covariance based methods is using a moving window of fixed size to determine the number of latest data collection points to include in the calculation, as first thought of by Bates and Granger in [22] and Granger and Ramanathan in [24]. Structural breaks can degrade the performance of these approaches. In [37], Pesaran uses a varying window size following a known structural break by minimizing the expected mean forecast error in an iterative procedure. Deutsch et al apply regime switching models to combine forecasts for the US and UK inflation rates in [38]. The regime the economy is in is then determined dependent on different functions of the lagged forecast error. A more computationally complex approach is followed in [39] and [40], where time varying coefficients are determined by applying a Kalman filter using an expanding window of observations.

Two of the fuzzy systems mentioned in the previous section include a learning mechanism: The authors of [32] periodically adapt the rule bases by calculating the confidence of an individual forecast based on their past performance. In [31], membership functions are automatically generated by processing incoming data, and thus adapted if new data arrives.

### 3.4 Discussion

Although there is a vast amount of literature available on linear forecast combination, no straightforward method of choosing the right approach has been found. The relative performance of the models depends on the error variance of the individual forecasts, the correlation between forecast errors and the sample size for estimation ([41]). Rank-based approaches work well for small sample sizes, while variance-covariance and regression methods are more suitable for bigger data sets. Similar error variances of individual forecasts indicate that simple averages might be a good choice.

Compared to literature about linear forecast combination, the number of publications about nonlinear methods appears small. Empirical results are mostly only smaller case studies, only one comparative study investigating neural networks, fuzzy systems and neuro-fuzzy systems for forecast combination has been found in [42]. Comparing algorithms using four different error measures, it is concluded that each of the presented nonlinear methods always outperforms both of the individual forecasts. The neuro-fuzzy approach performs best among the three.

Some publications about forecast combination with neural networks report a significant improvement over individual forecasting and linear forecast combination models ([43], [27]), however, in these two publications only one very short time series is used for evaluation. No details or justifications for the architecture of the neural network are given. A more extensive study finds that the performance of neural networks is mixed dependent on forecasting horizons or the error measure used ([28]). Self-organising algorithms have technical advantages compared to neural networks: They give an explicit model for each problem and the number of hidden layers and neurons does not have to be determined in advance. However, they seem to be a lesser-known and popular method in forecast combination with only one identified publication using it in a very small empirical study ([33]).

Results reported with time varying parameters are ambiguous. While Deutsch et al. ([44]) report significant improvements of the out-of-sample forecast error on one specific time series, more extensive empirical studies ([39], [40]) give discouraging results. According to these papers, estimating parameters with a recursive Kalman Filter approach seems to be the least fruitful approach, requiring relatively high computational power and yielding only small improvements. Switching regimes seem to be more promising.

Summarising, the combination of forecasts is generally considered as beneficial in most of the literature investigated ([18], [41]). When it comes to nonlinear forecast combination or combination with changing weights, the most optimistic studies are based on very small data sets while more extensive studies report mixed results, which indicates that the nonlinearity and adaptivity investigated is not beneficial for every time series.

## 4 Forecast and Forecast Combination - Experiments

A number of forecast and forecast combination algorithms have been described in the previous sections. Many of the empirical studies mentioned have one thing in common: forecasters spent a lot of time and apply a lot of knowledge in designing more or less sophisticated methods tuned for specific time series. In some bigger practical applications however, this is often not feasible, because often a great number of forecasts has to be calculated every second. Additionally, experts with sufficient knowledge to analyse a time series and choose an appropriate model for it are rare. This section describes an empirical experiment comparing the performance of off-the-shelf forecasting and forecast combination methods that have not been heavily tuned and are likely to be employed by users who are not forecasting experts.

### 4.1 Methodology

A data set consisting of eleven monthly empirical business time series with 118 to 126 observations each has been obtained from the 2006/2007 Forecasting Competition for Neural Networks and Computational Intelligence NN3 ([45]). Eighty percent of the data is used for training the forecasting models, the remaining twenty percent for assessing their out-of-sample performance. For the forecast combinations, those twenty percent have been divided into two equal parts, of which the first one is used for calculating combination weights and the second one for comparing out-of sample performances. The comparison between forecasts and forecast combinations can thus only take place in the last ten percent of the data set.

The mean squared error (MSE) and the mean absolute percentage error (MAPE) have been used for assessing and comparing methods. The MSE is used as an error measure in the majority of empirical forecast evaluations, while the MAPE as a relative error measure enables comparisons between different series.

The following forecasting methods have been implemented in Matlab:

1. The *naïve last observation* method, where the forecast is the latest observation.
2. The *moving average*, where the forecast is calculated according to the formula (2-1). A suitable size for the time window is found by calculating moving averages for values from 1 to 20 and choosing the value producing the lowest in-sample mean squared error.
3. *Single exponential smoothing* using formula (2-2), where again a grid search between 0 and 1 is employed to find the smoothing parameter  that produces the best in-sample MSE.
4. An *exponential smoothing approach with dampened trend* introduced by Taylor in [46] according to the formulas (4-1), where $L_t$ denotes the estimated level of the series and $R_t$ a growth rate. The smoothing parameters $\alpha$ and $\beta$ as well as parameter $\phi$ are again determined by a grid search.

$$\begin{aligned} \hat{y}_{t+1} &= L_t R_t^{\phi} \\ L_t &= \alpha y_t + (1-\alpha)(L_{t-1} R_{t-1}^{\phi}) \\ R_t &= \beta(L_t / L_{t-1}) + (1-\beta) R_{t-1}^{\phi} \end{aligned} \qquad (4\text{-}1)$$

5. A *polynomial regression*, where a polynomial of the order four is fitted to the time series by regressing time series indices against time series values.
6. The *theta-model* according to formula (2-4) using formulas given in [47], where two curves are used and eventually combined by a simple average. One curve

originates from $\phi = 0$, yielding a line that can be easily extrapolated; the other curve uses $\phi = 2$ and predicts its future values by single exponential smoothing.

7. An *ARIMA* model following section 2.4. Since it is difficult to identify an appropriate ARIMA model automatically, a pragmatic approach was used: the first difference of the series has been taken to remove a certain amount of non-stationarity before submitting it to an automatic ARMA selection process ([48]).
8. A *Feed-forward neural network*, whose 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 [17]. According to this publication, a backpropagation algorithm with momentum is successfully used by the majority of empirical studies. One hidden layer is generally considered sufficient for forecasting purposes. It is suggested to use as many hidden neurons as input variables, with input variables being chosen with regard to the application. The number of lagged input variables and thus the number of hidden neurons has been set to 12 for the experiments here in order to capture yearly seasonal effects, if present. A grid search of the nine possible variations of the values 0.1, 0.5 and 0.9 for the learning rate and the value of the momentum has been carried out, indicating best results with a learning rate of 0.1 and a momentum of 0.5.

Furthermore, the following forecast combination methods have been employed:

1. The *simple average*, which calculates the average of all available forecasts.
2. Similar to the previous method, *simple average with trimming* averages individual forecasts, but only the best performing 80% of the models are taken into account. Performance was assessed by the MSE of the validation period.
3. A variance-based forecast combination method using formula (3-3).
4. The traditional *outperformance* method of Bunn, which was mentioned in section 3.1. Past performance is again assessed in the validation period.
5. A *variance-based pooling approach* introduced by Aiolfi and Timmermann ([35]). Forecasts are grouped into two clusters by a k-means algorithm according to the mean squared error of the validation period. Forecasts of the historically best performing cluster are then simply averaged to obtain a final forecast.
6. A second variance-based pooling method, which is almost identical to the previous one. The only difference is that three clusters are used rather than two.

### 4.2 Results

Table 4.1 shows average out-of-sample MSE and MAPE values for the different forecasting methods numbered from one to ten. MSE values are given both relative to the naive last observation forecast and in absolute values. It has to be noted that the average absolute MSE values have limited explanatory power and are only given for illustrative purposes, because the scale of each time series is different. The most obvious result is the superior performance of the neural network (method eight), strongly dominating average MAPE values, reducing the next best method's MAPE by further 41%. Using the relative MSE measure, the neural network performs marginally worse than methods three, four and six. The ARMA method, which represents the most statistically complex method used in this test, fails to outperform any other method apart from the regression approach. Worth mentioning is Taylor's modified dampened trend exponential smoothing (method number four), which is very easy to use and ranks best using the MSE and second best using the MAPE error measure.

**Table 4.1.** MSE and MAPE error values for forecasting methods, measured on the 20% testing set and averaged for each of the 8 methods over the 11 series.

| **Method** | **MSE (relative to naïve forecast)** | **MSE (absolute, x $10^7$)** | **MAPE** |
|---|---|---|---|
| 1 | 1.00 | 2.7524 | 55.86 |
| 2 | 0.91 | 2.6886 | 55.14 |
| 3 | 0.88 | 2.6988 | 54.06 |
| 4 | 0.86 | 2.5561 | 28.31 |
| 5 | 98.32 | 27.972 | 3473.77 |
| 6 | 0.88 | 2.6988 | 55.34 |
| 7 | 1.40 | 2.6793 | 90.80 |
| 8 | 0.90 | 1.1682 | 16.67 |

Looking at the performance of forecast combination methods in table 4.2, variance-based pooling with three clusters (method number six) is the best performing approach in terms of the average MAPE error measure. The MSE gives different average results: one big outlier for one of the time series in the data set corrupts this method's overall performance, the variance-based method number three performs best instead. Simple average with trimming results in second best forecasts for both error measures.

**Table 4.2.** MSE and MAPE error values for forecast combination methods, measured on the 10% testing set and averaged for each of the 6 methods over the 11 series.

| **Method** | **MSE (relative to naïve forecast)** | **MSE (absolute, x $10^7$)** | **MAPE** |
|---|---|---|---|
| 1 | 3.53 | 2.7229 | 57.36 |
| 2 | 0.79 | 2.1934 | 23.36 |
| 3 | 0.72 | 2.0632 | 26.85 |
| 4 | 8.26 | 1.6656 | 21.93 |
| 5 | 0.87 | 2.2638 | 24.08 |
| 6 | 1.66 | 0.96169 | 16.68 |

Table 4.3 compares forecasts to forecast combinations by showing the number of the best performing method according to the MSE and MAPE of the validation period together with the corresponding MSE and MAPE values of the testing period. Forecast combinations are clearly beneficial in this experiment for both error measures; for four out of the eleven time series they outperform individual forecast algorithms, five times they perform equally. In turn, they are only outperformed on two of the series. Regarding the MAPE error measure, the error values on the two series where the combinations performed worse are still quite similar, while on the other hand, improvements of up to 12% could be achieved from using them on other series. This does not hold looking at the MSE, where differences are more significant – using

combinations results in improvements of up to 40% (series eight), but also in deteriorating performance of up to 72% (series one).

**Table 4.3.** Best performing method chosen in validation period and corresponding MSE (relative to naïve forecast) and MAPE error measures of the testing period, left: forecasting methods, right: forecast combination methods.

| | Forecasting | | | Forecast Combination | | |
|---|---|---|---|---|---|---|
| **Series** | **method (MSE/ MAPE)** | **MSE (rel)** | **MAPE** | **method (MSE/ MAPE)** | **method (MSE/M APE)** | **MAPE** |
| 1 | 8/4 | 0.42 | 2.74 | 5/2 | 1.48 | 2.73 |
| 2 | 8/8 | 0.46 | 17.27 | 6/6 | 0.46 | 17.27 |
| 3 | 8/8 | 0.33 | 44.31 | 6/6 | 0.33 | 44.31 |
| 4 | 8/8 | 0.39 | 16.44 | 6/6 | 0.38 | 14.96 |
| 5 | 8/8 | 2.07 | 3.24 | 6/6 | 2.07 | 3.24 |
| 6 | 8/8 | 0.62 | 6.87 | 2/2 | 0.70 | 6.94 |
| 7 | 5/5 | 9.31 | 9.65 | 6/6 | 9.31 | 9.65 |
| 8 | 4/8 | 0.29 | 11.57 | 6/6 | 0.25 | 10.50 |
| 9 | 7/7 | 0.57 | 4.27 | 6/6 | 0.57 | 4.27 |
| 10 | 8/8 | 4.87 | 25.36 | 6/6 | 2.91 | 22.40 |
| 11 | 3/8 | 0.90 | 20.34 | 6/1 | 0.82 | 21.95 |
| avg | 8/8 | 1.13 | 14.13 | 6/6 | 1.66 | 14.06 |

The results can be summarised as follows: When nothing is known about a time series and an individual forecasting method is needed, Taylor’s dampened trend exponential smoothing is a safe option to improve upon the naïve last observation forecast. When only considering a MAPE loss-function, a neural network is likely to achieve consistent improvements too. Combinations of forecasting methods outperformed forecasts of individual approaches for the majority of the series. A simple average with trimming performed comparatively well for both the MSE and the MAPE error measure; variance-based pooling proved beneficial when working with MAPE error measures.

## 5 Conclusion and Future Work

This paper presented and discussed forecast and forecast combination approaches with a focus on their application in airline industry. Building on a short introduction to airline revenue management, the importance and need for effective forecasting procedures in airline industry has been explained and justified. Two different approaches to forecasting demand and cancellation rates that are crucial in this context have been identified, one involves traditional time series forecast and forecast combination methods further described in sections two and three. The second approach is prediction using data on passenger level, which is an area of research in airline industry forecasting that was made possible with the change from leg-based to O&D systems in the 1990s. Currently, related work only seems to have been done in the area of forecasting no-show rates. Testing its contributions for demand and cancellation rates is an interesting and novel issue.

An empirical evaluation on a publicly available data set has been carried out, demonstrating the potential of forecast combinations. However, the combination of forecasts was not equally successful for all series and the question why some approaches work well on most of the series while completely failing to reasonably predict others remains unanswered and is an area for future research. Forecast combinations have also been proven to be very successful in a previous project when used to forecast demand for airline tickets, their usefulness to predict cancellation rates has yet to be investigated.

## References


1. Riedel, S., Gabrys, B., (2004). *Hierarchical Multilevel Approaches of Forecast Combination*, Proceedings of the GOR 2004 conference, p. 1-3.
2. Riedel, S., Gabrys, B., (2005). *Evolving Multilevel Forecast Combination Models-An Experimental Study,* Proceedings of NiSIS 2005 Symposium.
3. Riedel, S., Gabrys, B., (2007). *Dynamic Pooling for the Combination of Forecasts generated using Multi Level Learning,* Proc.of the International Joint Conference on Neural Networks.
4. McGill, J.I., van Ryzin, G.J., (1999). *Revenue management: Research overview and prospects,* Transportation Science, **33**(2), p. 233-256.
5. Makridakis, S.G., Wheelwright, S.C., Hyndman, R.J., (1998). *Forecasting: Methods and Applications,* 3rd ed., New York, John Wiley.
6. Clemen, R.T., (1989). *Combining forecasts: A review and annotated bibliography,* International Journal of Forecasting, **5** p. 559-583.
7. Timmermann, A., (2006). *Forecast Combinations*, in *Handbook of Economic Forecasting*, North Holland, Elsevier, p. 135-196.
8. Baesens, B., et al., (2004). *Bayesian network classifiers for identifying the slope of the customer lifecycle of long-life customers,* European Journal of Operational Research, **156**(2), p. 508-523.
9. Maes, S., et al., (2002). *Credit Card Fraud Detection Using Bayesian and Neural Networks,* Proceedings of the 1st International NAISO Congress on Neuro Fuzzy Technologies.
10. Giudici, P., Passerone, G., (2002). *Data mining of association structures to model consumer behaviour*, Computational Statistics and Data Analysis, **38**(4), p. 533-541.
11. Neuling, R., Riedel, S., Kalka, K.U., (2004). *New approaches to origin and destination and no-show forecasting: Excavating the passenger name records treasure*, Journal of Revenue and Pricing Management, **3**(1).
12. Lawrence, R.D., Hong, S.J., Cherrier, J., (2003). *Passenger-based predictive modeling of airline no-show rates,* Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, p. 397-406.
13. Gardner, E.S., (1985). *Exponential Smoothing: The State of the Art*, Journal of Forecasting, **4**(1), p. 1-28.
14. Assimakopoulos, V., Nikolopoulos, K., (2000). *The theta model: a decomposition approach to forecasting,* International Journal of Forecasting, **16**(4), p. 521-30.
15. Box, G.E.P., Jenkins, G.M., (1970). *Time series analysis,* San Francisco, Holden-Day.

16. van Dijk, D., Teräsvirta, T., Franses, P.H., (2002). *Smooth Transition Autoregressive Models: A Survey of Recent Developments,* Econometric Reviews, **21**(1), p. 1-47.
17. Zhang, G., Patuwo, B.E., Hu, M.Y., (1998). *Forecasting with artificial neural networks: The state of the art,* International Journal of Forecasting, **14**(1), p. 35-62.
18. Makridakis, S., Hibon, M., (2000). *The M3-Competition: results, conclusions and implications,* International Journal of Forecasting, **16**(4), p. 451-476.
19. Ord, K., (2001). *Commentaries on the M3-Competition,* International Journal of Forecasting, **16**(4), p. 537-584.
20. Stock, J.H., Watson, M.W., (1998). *A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series*, National Bureau of Economic Research.
21. Teräsvirta, T., van Dijk, D., Medeiros, M.C., (2004). *Linear Models, Smooth Transition Autoregressions, and Neural Networks for Forecasting Macroeconomic Time Series: A Reexamination,* Pontifícia Universidade Católica de Rio de Janeiro, Departamento de Economía.
22. Bates, J., Granger, C., (1969). *The Combination of Forecasts,* Operational Research, **20**(4), p. 451-468.
23. Newbold, P., Granger, C., (1974). *Experience with Forecasting Univariate Time Series and the Combination of Forecasts,* Journal of the Royal Statistical Society. Series A (General), **137**(2), p. 131-165.
24. Granger, C.W.J., Ramanathan, R., (1984). *Improved methods of combining forecasts,* Journal of Forecasting, **3**(2), p. 197-204.
25. Bunn, D.W., (1975). *A Bayesian Approach to the Linear Combination of Forecasts,* Operational Research Quarterly, **26**(2), p. 325-329.
26. Gupta, S., Wilton, P.C., (1987). *Combination of Forecasts: An Extension,* Management Science, **33**(3), p. 356-372.
27. Shi, S., Da Xu, L., Liu, B., (1999). *Improving the accuracy of nonlinear combined forecasting using neural networks,* Exp. Systems with Applications, **16**(1), p. 49-54.
28. Donaldson, R.G., Kamstra, M., (1996). *Forecast combining with neural networks*, Journal of Forecasting, **15**(1), p. 49-61.
29. Hu, M., Tsoukalas, C., (1999). *Combining conditional volatility forecasts using neural networks: an application to the EMS exchange rates,* Journal of International Financial Markets, Institutions & Money, **9**(4), p. 407-422.
30. Fromm, J., Fei, H., Fiordaliso, A., (1998). *A nonlinear forecasts combination method based on Takagi-Sugeno fuzzy systems,* International Journal of Forecasting, **14**(3), p. 367-379.
31. Frantti, T., Mähönen, P., (2001). *Fuzzy logic-based forecasting model,* Engineering Applications of Artificial Intelligence, **14**(2), p. 189-201.
32. Petrovic, D., Xie, Y., Burnham, K., (2006). *Fuzzy decision support system for demand forecasting with a learning mechanism,* Fuzzy Sets and Systems, **157** p. 1713-1725.
33. He, C., Xu, X., (2005). *Combination of Forecasts Using Self-organizing Algorithms*, Journal of Forecasting, **24** p. 269-278.
34. Ivakhnenko, A., (1970). *Heuristic self-organization in problems of engineering cybernetics,* Automatica, **6**(2), p. 207-219.

35. Aiolfi, M., Timmermann, A., (2006). *Persistence in forecasting performance and conditional combination strategies*, Journal of Econometrics, **127**(1-2), p. 31-53.
36. Diebold, F.X., Pauly, P., (1986). *Structural change and the combination of forecasts*, Board of Governors of the Federal Reserve System (U.S.).
37. Pesaran, M., Timmermann, A., (2007). *Selection of Estimation Window in the Presence of Breaks*, Journal of Econometrics, **127**(1), p. 134-161.
38. Deutsch, M., Granger, C.W.J., Diebold, F.X., (1988). *Serial Correlation and the Combination of Forecasts,* Journal of Business & Economic Statistics, **6**(1), p. 105-111.
39. Terui, N., van Dijk, H.K., (2002). *Combined forecasts from linear and nonlinear time series models,* International Journal of Forecasting, Volume 18, Issue 3, p. 421-438.
40. Stock, J.H., Watson, M.W., (2004). *Combination forecasts of output growth in a seven-country data set*, Journal of Forecasting, **23**(6), p. 405-430.
41. de Menezes, L.M., Bunn, D.W., Taylor, J.W., (2000). *Review of guidelines for the use of combined forecasts,* European Journal of Operational Research, **120**(1), p. 190-204.
42. Palit, A.K., Popovic, D., (2000). *Nonlinear combination of forecasts using artificial neural network, fuzzy logic and neuro-fuzzy approaches,* Proc.of the ninth IEEE International Conference on Fuzzy Systems, p. 566-571.
43. Shi, S., Liu, B., (1993). *Nonlinear combination of forecasts with neural networks,* Proc.of 1993 International Joint Conference on Neural Networks.
44. Deutsch, M., Granger, C.W.J., Teräsvirta, T., (1994). *The combination of forecasts using changing weights,* International Journal of Forecasting, **10**(1), p. 47-57.
45. *NN3 Artificial Neural Network and Computational Intelligence Forecasting Competition [Online]*. [cited 2007 13/06], available from: http://www.neural-forecasting-competition.com/
46. Taylor, J.W., (2003). *Exponential smoothing with a damped multiplicative trend,* International Journal of Forecasting, **19**(4), p. 715-725.
47. Hyndman, R.J., Billah, B., (2003). *Unmasking the Theta method,* International Journal of Forecasting, **19**(2), p. 287-290.
48. *Delft Center for Systems and Control - Software*. [Online][cited 2007 13/06], available from: http://www.dcsc.tudelft.nl/Research/Software