Air Travel Itinerary Market Share Estimation

Tobias Grosche, Franz Rothlauf

Working Paper 6/2007
March 2007

Working Papers in Information Systems

University of Mannheim
Department of Information Systems
D-68131 Mannheim/Germany
Phone +49 621 1811691, Fax +49 621 1811692
E-Mail: wifo1@uni-mannheim.de
Internet: http://wifo1.bwl.uni-mannheim.de

Air Travel Itinerary Market Share Estimation

Tobias Grosche *, Franz Rothlauf

Department of Information Systems I, University of Mannheim, D-68131 Mannheim

Abstract

In air travel, an itinerary is a direct flight or sequence of connecting flights between two cities. The objective of itinerary market share estimation is to forecast market shares of competing itineraries. This paper examines and compares three different methods for itinerary market share estimation: multinomial logit models, artificial neural networks, and a custom model developed by the authors. Using real-world booking data, each model is constructed and calibrated to best reproduce the given data. The resulting models are applied to test data and the custom model was found to show the best results. Although multinomial logit model are used by many airlines for planning and forecasting purposes, such methods resulted in the lowest forecasting quality.

* Corresponding author.
Address:
Department of Information Systems I
University of Mannheim
D-68131 Mannheim
Germany
Tel: +49 (0) 621 / 181 - 1691
Fax: +49 (0) 621 / 181 - 1692
E-Mail: tobias.grosche@uni-mannheim.de

1 Introduction

A major task in airline planning and scheduling is forecasting. Decisions on resource allocation, flight scheduling, and pricing rely on different types of forecasts, reaching from long-term strategic forecasts of travel demand between regions/countries to short-term forecasts of no-show and go-show passengers just prior to a departing flight. An itinerary is either a direct flight or a sequence of connecting flights between two cities. Forecasts on the itinerary level are important for accurate planning and flight schedule construction since itineraries are the products that are finally purchased by the passengers (Coldren and Koppelman, 2005). Although there are a number of publications on forecasting techniques and studies of passenger demand forecasting (see Coldren and Koppelman (2005) for an overview), few published models are available which are able to forecast travel demand for itineraries.

The objective of itinerary market share estimation is to forecast the number of passengers that are expected to book an itinerary. The itinerary's share is the number of passengers divided by the total number of passengers on the same day and route (market). This share can be interpreted as the attraction of an itinerary for a single passenger. It depends on attributes such as convenience of travel, travel time, departure and arrival time, average fare, aircraft type, and airline preferences.

In this paper, three different methods for itinerary market share estimation are studied and compared: a multinomial logit (MNL) model, an artificial neural network (ANN), and a custom model for the estimation of itinerary shares (EIS). For each method, one model is constructed, calibrated, and evaluated using historical booking data. Because the input data are the same for all models, they can be compared well and evaluated according to their forecasting quality.

This paper is organized as follows: the next section describes the basic setup including the booking data and approach used for calibration and evaluation of all models. Sections 3, 4, and 5 present the MNL model, the ANN model, and the EIS model, respectively. These sections also give details on the calibration and evaluation of the models. In section 6 the forecasting quality of all three models is compared.

2 Basic Setup

2.1 Overview

An important task in constructing forecasting models is choosing an appropriate model structure. The model structure also includes method parameters (for example, coefficient of a regression model) which are usually calibrated by using given data of the problem to maximize prediction quality of the model. For itinerary share estimation, the method parameters describe the impact of the attributes of the itineraries (independent variables) on its attractiveness (dependent variable). Thus, historical data used for calibration must measure the realized passenger demand (as a share of the total demand in the city pair) and the corresponding attributes of the itineraries.

The impact of attributes of itineraries can be modeled either separately for each city pair, or aggregated for all city pairs. If modeled separately for each city pair, model parameters are different between markets, whereas the aggregated calibration - as conducted in this paper - results in method parameters applicable to all city pairs. Thus, the model can be used for estimation in new markets which is important for flight schedule construction. In addition, this study does not use different passenger segments or time periods resulting in group-specific or time-specific coefficients. Instead, each model is calibrated using all available data.

2.2 Data

In this study, MIDT [1] booking data from January to August 2004 for itineraries between Germany and European countries was used. The booking data contained direct flights and connections with a maximum of one stop. Only markets with at least two itineraries are considered for the study (if only one itinerary exists no estimation of the market share is necessary). The resulting data set contained 2,978 different city pairs with a total of 961,430 itineraries, and a total number of passengers on these itineraries of 7,312,610.

2.2.1 Input Variables

In principle, the number of attributes of an itinerary can be large depending on the level of detail. Table 1 lists the attributes (independent variables) that are used for this study to describe relevant properties of itineraries. It also

[1] Market Itinerary Data Tapes

presents a short description of each variable, its range, and if necessary, the functional form as used in the different models. The different variables are modeled such that the impact of the variable on the attraction of an itinerary increases with higher values.

variable	values	functional form	description
travel time ratio	[0,1]	$TTR_i = \max(2 - \frac{time_i}{time_{sh}}, 0)$	Ratio between total travel time $time_i$ of itinerary i and travel time $time_{sh}$ of shortest itinerary sh in the market.
itinerary type	{0,1}	$TYP_i = \begin{cases} 1 & \text{if } i \text{ is direct flight} \\ 0 & \text{if } i \text{ is connection} \end{cases}$	Discrete value indicating direct flight or connection.
shortest itinerary type	{0,1}	$STY_i = \begin{cases} 1 & \text{if } sh \text{ is connection} \\ 0 & \text{if } sh \text{ is direct flight} \end{cases}$	Discrete value indicating if shortest itinerary sh in the market is direct flight or connection.
departure time preference	[0,1]	$DTP(dep_i)$	Indicates the attraction of the departure time dep_i of itinerary i for a potential passenger (see Figure 1).
airline quality/ preference	[0,1]	QUA_i	Describes the quality of the airline operating itinerary i as published in Skytrax (2006).
airline presence	[0,1]	PRS_i	Indicates the total market share of the airline operating itinerary i in the market.
closeness (closest itinerary)	[0,144]	$CLO = \begin{cases} 0 & \text{if } i = cl \\ 144 - \lvert dep_i - dep_{cl} \rvert & \text{else} \end{cases}$	Time difference between departure time dep_i of itinerary i and departure time dep_{cl} of the closest (with respect to time) itinerary in the market. Time is measured in 5-minute-intervals (maximal time difference is 144 (12 hours)).
travel time ratio (closest itinerary)	[0,2]	$TRC_i = 2 - \frac{time_i}{time_{cl}}$	Ratio between total travel time $time_i$ of itinerary i in comparison to travel time $time_{cl}$ of the closest itinerary in the market.

Table 1
Description of explanatory variables

The variable $DTP(t)$ requires further explanation as time preferences do not stay constant during the day because travelers usually have preferences for specific departure times. For example, standard business travelers are likely to prefer departure times in the morning and in the afternoon/evening. $DTP(t)$ describes how the preference for a specific departure time changes throughout a day. In this study, three different $DTP(t)$ functions are considered:

- USA70: This function is derived from a survey of domestic airline traffic conducted in 1969 by the US Department of Transportation (O'Connor, 1982).
- AXS: This function is used in software used by an airline for schedule evaluation.
- EU86: This function is derived from a study in 1986 on passenger volumes on short-haul routes in Europe published by Biermann (1986).

Figure 1 plots the three different functions.

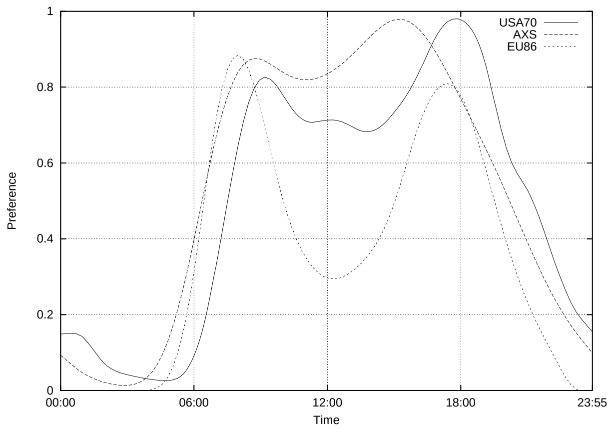


Fig. 1. Three different functions for $DTP(t)$

2.2.2 Output Variables

Because it is difficult to obtain correct values for the unconstrained demand, historical booking data is used as a substitute of the demand. Consequently, the total demand in one market is calculated as the sum of all bookings over all itineraries in the market. By dividing the number of passengers on one itinerary by the total number of passengers in the market, the market share of this itinerary can be calculated. By using the market share as the dependent variable, we are able to build an aggregate forecasting model for all city pairs and eliminate the effects of different market sizes.

2.3 Calibration and Evaluation

The goal of calibration is to adjust the method parameters of a forecasting model such that the model reproduces well the calibration data. In this study, the process of calibration and evaluation is the same for all three models. Out of the total number of observations, a set of randomly chosen itineraries serves either as a calibration data set (CS) or as a validation data set (VS). By using the CS, each model is calibrated until no further improvement of the forecasting quality is possible. Then, the calibrated model is evaluated by measuring the forecasting quality using the data of the VS. The forecasting quality of each model is evaluated using the mean squared error

$$MSE = \frac{\sum_k (p_k - t_k)^2}{|K|},$$

where $|K|$ is the number of elements in the total set K of itineraries, p_k is the market share predicted for itinerary $k \in K$, and t_k is the observed market share.

3 Multinomial Logit Model

3.1 Overview

In this section, a multinomial logit (MNL) model for itinerary market share forecasting is formulated and tested. Multinomial logit models are commonly used methods for *Discrete Choice Problems* in which a person has to choose one alternative from a given (finite discrete, explicit listed) set of alternatives. Although MNL models are common in marketing research and airline planning, only a few publications of MNL models for itinerary market share estimation are available (Coldren and Koppelman, 2005). For examples see Ashford and Benchemam (1987), Alamdari and Black (1992), Coldren et al. (2003), and Hsu and Wen (2003).

The following section formulates the MNL model for the given problem. In section 3.3 the given data is used in the MNL model for the calibration and validation process.

In this section, a MNL model for the itinerary market share estimation problem is presented. See Train (2003), Ben-Akiva and Lerman (1985), or Kanafani (1983) for more details.

In general, it is assumed that each passenger acts rationally and wants to maximize his utility (Ben-Akiva and Bierlaire, 1999). The utility or value V_k of a given itinerary $k \in K$ for a passenger $n \in N$ depends on the attributes $a \in A$ of the itinerary. In MNL models, V_k is a linear combination of the attributes values $\mathbf{X_k} = (x_{k1}, x_{k2}, \ldots, x_{ka})$ and method parameters $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_a)$:

$$\begin{aligned} V_k &= \boldsymbol{\beta}^T \mathbf{X_k} \\ &= \beta_1 x_{k1} + \beta_2 x_{k2} + \beta_3 x_{k3} \ldots \beta_a x_{ka}. \end{aligned}$$

In this formulation, the value of one itinerary depends only on the characteristics of this alternative. Attributes of individual passengers are not included in the model and all passengers are grouped to one single entity with the same value perception. Furthermore, we assume one vector $\boldsymbol{\beta}$ for the total model, and do not build individual parameters for segments of the total number of observations.

The probability p_k (attraction) of an itinerary k to be chosen by one passenger $n \in N$ is defined by:

$$p_k = \frac{e^{V_k}}{\sum\limits_k e^{V_k}}.$$

During calibration, the objective is to determine the vector $\hat{\boldsymbol{\beta}}$ that maximizes the likelihood of the observation. We define y_{nk} as:

$$y_{nk} = \begin{cases} 1 & \text{if individual } n \text{ chooses alternative } k \\ 0 & \text{otherwise.} \end{cases}$$

Then, the probability of the correct choice of individual n is given by

$$\prod_k (p_k)^{y_{nk}}.$$

Because $y_{nk} = 0$ for all non-chosen alternatives, this term is simply the probability p_k of the chosen alternative k.

Since the individual choices are independent, the probability of the correct prediction of all N individual choices is given by the likelihood function:

$$L(\beta) = \prod_{n=1}^{N} \prod_{k} (p_k)^{y_{nk}}.$$

We define D_k as the number of passengers that choose alternative k as:

$$D_k = \sum_{n=1}^{N} y_{nk}$$

and D as the total number of passengers:

$$D = \sum_{k} D_k.$$

Because the choices of different individuals are assumed to be independent and identically distributed (Bernoulli trials), the joint probability is given by the multinomial distribution. Therefore, the likelihood function can be calculated as:

$$L(\beta) = \frac{D!}{D_1!D_2!\dots D_k!} \prod_{k} (p_k)^{D_k}.$$

This likelihood function is for a given set of competing itineraries. When considering different city pairs and days in model calibration, itineraries have to be separated into individual groups with competition within but not between. By defining a market as the combination of a city pair and a day, M as the total set of markets, $m \in M$ as one market, and K_m as the set of itineraries competing in market m, the likelihood function is given as:

$$L(\beta) = \prod_{m=1}^{M} \frac{D_m!}{\prod\limits_{k \in K_m} D_k!} \prod_{k \in K_m} (p_k)^{D_k}.$$

The maximum likelihood estimator (MLE) is the value $\hat{\boldsymbol{\beta}}$ that maximizes this function. Taking the logarithms simplifies maximization resulting in the log-likelihood function:

$$LL(\beta) = \sum_{m=1}^{M} \left\{ \ln D_m! - \sum_{k \in K_m} \ln D_k! + \sum_{k \in K_m} D_k \ln p_k \right\}.$$

$LL(\beta)$ is globally concave with respect to $\boldsymbol{\beta}$ simplifying the estimation of $\hat{\boldsymbol{\beta}}$ (McFadden, 1974).

To test the overall model (structure), the log-likelihood ratio index ρ^2 is computed as:

$$\rho^2 = 1 - \frac{LL(\hat{\boldsymbol{\beta}})}{LL(0)}$$

with $LL(\hat{\boldsymbol{\beta}})$ as the value of the log-likelihood function for the estimated value $\hat{\boldsymbol{\beta}}$ and $LL(0)$ as the log-likelihood value for $\boldsymbol{\beta} = 0$. Because additional independent variables never reduce ρ^2, the log-likelihood ratio is corrected by the number of variables A:

$$\bar{\rho^2} = 1 - \frac{LL(\hat{\boldsymbol{\beta}}) - A}{LL(0)}.$$

Although having the same objective as the coefficient of determination R^2 in classical linear regression, the interpretation of $\bar{\rho^2}$ is not exactly the same. Values of $\bar{\rho^2}$ between 0.2 and 0.4 are assumed to indicate an acceptable model fit (Urban, 1993).

To determine the significance of individual variables, standard t-tests can be used (Train, 2003). For each β_a the null hypothesis H_0 is tested against the alternative hypothesis H_α:

$$H_0 : \beta_a = 0$$
$$H_\alpha : \beta_a \neq 0.$$

3.3 Calibration and Validation

In model calibration, the objective is to find $\hat{\boldsymbol{\beta}}$ which is conducted by Maximum-Likelihood-Estimation. All variables described in Table 1 were included, however, there is a choice between the three different time preference functions. In a set of experiments, the impact of the different time preferences is analyzed. Using each time preference function separately, an MNL model is calibrated and validated in 10 experiments, each with a different set of 40,000 randomly chosen itineraries as CS and VS. Figure 2 shows the averages of the LL-Ratio and MSE (of the VS) for each time preference function is illustrated.

The consideration of the different time preference functions yield to similar results, and the LL-Ratio between 0.313 and 0.317 indicate an acceptable model fit. Because USA70 has the lowest MSE, this time preference function should be used within this MNL model for itinerary share estimation.

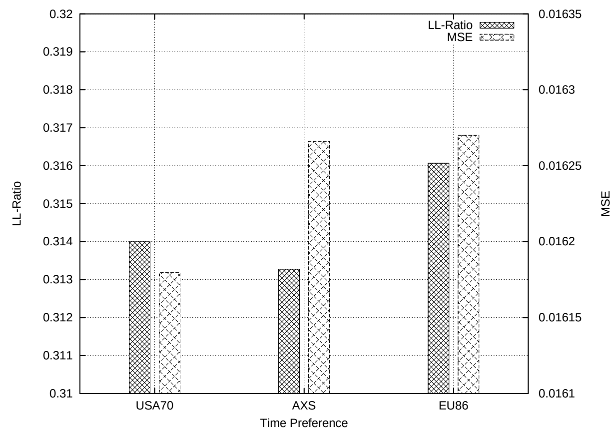


Fig. 2. Time preference

4 Artificial Neural Network

4.1 Overview

The main reason why MNL models are commonly used in forecasting are their well-defined structure leading to easy calibration and fast computation times. However, it remains unclear whether the basic structure of MNL models (logistic function, linear-in-parameter utility) limits the forecasting accuracy in comparison to other structures or less structured models. One approach used in forecasting that does not exogenously provide a given functional relationship between attributes of a possible choice and its selection probability are artificial neural networks (ANN).

For ANNs, the calibration process can be described as a trial-and-error process resulting in a model structure which is difficult to be interpreted. The relationship between input and output variables can not be interpreted as in MNL models, where statements regarding the relative importance or individual impact of certain attributes on the output are possible. However, if the goal is to accurately predict passenger behavior and there is no interest in the functional relationship between input and output variables, ANNs are an interesting alternative to MNL models. Because ANN have no pre-determined model structure, they might increase forecasting quality at the cost of a more difficult and time-consuming calibration process.

Until now, ANN have not yet be applied to the itinerary market share problem. Weatherford et al. (2003) and Nam et al. (1997) used ANNs to forecast air passengers on a more general level for different fare classes and days of the week, outperforming traditional forecasting techniques (like moving averages, exponential smoothing, regression, and others).

In the next section, a short introduction to ANN is given and the process of model construction and calibration is illustrated. Section 4.3.1 describes the calibration of ANNs for the given problem. For all experiments in this study *SNNS Stuttgart Neural Network Simulator 4.1* was used (Zell et al., 1995).

4.2 Underlying Principles

ANN imitate biological information processing on an abstract level. They consist of artificial neurons that are simple information processing units and links connecting the different neurons (Arbib, 2002). ANNs exhibit complex global behavior by arranging the neurons in interconnected groups (serial, parallel). In a training phase, an ANN can change its structure based on a given learning function to best fit the information or data.

Feed-forward networks are a common ANN type. Here, different layers of neurons exist and each neuron in one layer receives the output of neurons of the previous layer. Usually, the input variables of a problem are represented as an input layer of neurons, the dependent variables as an output layer. Depending upon the complexity of the problem, several hidden layers connecting input and output layer may be necessary (Weatherford et al., 2003). Figure 3 gives an example with four input neurons, two hidden layers (with 5 and 3 neurons, respectively), and one output neuron.

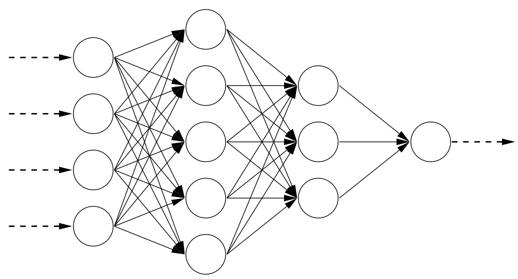


Fig. 3. Example of a multi layer neural network

Usually, weights are assigned to links between neurons to modify the information transfered to the next neuron. These weights are adjusted in a training phase such that the training data is correctly predicted (Hruschka, 1993). In the training phase, all training data (consisting of values for the input and output variables) is iteratively used to adjust each weight until the prediction error of the ANN is below a certain threshold. After finishing the training phase, the ANN can be used to forecast values for new problem instances.

In addition to selecting a learning function, a proper network topology has to be selected and the information processing steps performed in each neuron (neuron functions) have to be specified. The topology of a network can vary in the number of layers, the number of neurons in each layer, and the connectivity between the layers. Neuron functions usually consist of a propagation function, an activation function, and an output function. The propagation function f_j^{net} describes how the input value net_j of a neuron n_j depends on the outputs out_i of all previous neurons n_i and the weights w_{ij} between n_i and n_j. Often, the weighted sum $net_j = \sum_i w_{ij} out_i$ is used. The activation function calculates the neurons activation state $act_j = f_j^{act}(net_j)$, and the output function $out_j = f_j^{out}(act_j)$ determines the output value out_j of n_j. In most ANN applications, including this study, the output is equivalent to the activation state ($out_j = act_j$).

Although the training process and application of ANNs is simple, finding proper ANN structures for a given problem is a complex design task. There is no standard method or procedure on how to select a network configuration consisting of the learning function (and its parameters), the topology, and the neuron functions. Thus, developing a proper ANN design is usually a time-consuming trial-and-error-process in which different configurations are tested and evaluated.

4.3 Calibration and Validation

4.3.1 Overview

Before ANNS can be used for passenger estimation, a proper ANN design must be determined which results in a high forecasting quality. Because no algorithm or standard procedure exists for this task, usually many different ANN designs are tested and compared. In this study, we focus on the specification of the following elements:

- Time preference function,
- learning function and parameters, and
- network topology (number of hidden layers and neurons).

The other elements of an ANN (such as input and output neurons and their neuron functions, termination criteria for learning) have been evaluated in preliminary tests not reported in this paper. These studies resulted in a basic configuration which is described in the next section.

4.3.2 Basic Configuration

The basic configuration includes some decisions on the neuron layers and functions and the ANN's learning:

- **Neuron layers and functions**

  Input Layer: The input layer contains one neuron n_i for each input variable shown in Table 1. In MNL and EIS models, the number of itineraries competing in one market is implicitly contained in the model structure. In contrast to this, in ANN the (normalized) number of competing itineraries has to be included as an additional input neuron. Because the input layer only processes the given input values x_i, the propagation, activation and output function are the identity function ($out_i = f_i^{out} = f_i^{act} = f_i^{net} = x_i$). To eliminate problems with different scales of input variables, CLO is normalized to $[0, 1]$.

  Hidden Layers: The number of neurons in each hidden layer and the number of hidden layers is determined by the topology and is described in Section 4.3.3. Preliminary tests showed good results when using the weighted sum as propagation function and the tangens hyperbolicus as activation function in the hidden neurons n_k (independently of the used topology). Therefore,

$$f_k^{net} = \sum_j w_{jk} out_j,$$
$$f_k^{act} = \frac{e^{net_k} - e^{-net_k}}{e^{net_k} + e^{-net_k}}.$$

  Output Layer: The output layer has only one neuron n_o representing the dependent variable (market share). We choose the logistic function as the activation function and the weighted sum as the propagation function as these produced good results in the preliminary tests:

$$f_o^{net} = \sum_k w_{ko} out_k,$$
$$f_o^{act} = \frac{1}{1 + e^{-net_o}}.$$

  Connectivity: The connectivity is determined by the topology and specifies which neurons are connected by a link. In this study, first order feed-forward networks are used. There are no direct links between neurons in the same layer and each neuron is directly connected to all neurons in the following layer.
- **Learning**

  In analogy to the calibration process for the MNL model, a calibration and validation set is used. The calibration set includes the training data that is used in the training phase to adjust the weights of the ANN such that the mean squared error is minimized. Afterwards, the weights are fixed and the ANN is applied to the validation set.

  During training the ANN learns to reproduce the training data but looses its ability to generalize. Thus, it is important to stop the training phase when the best compromise between generalization and prediction accuracy is reached. Figure 4 shows how the MSE depends on the number of epochs (application of the complete set of training data) for an example training phase of the ANN. After each epoch the MSE is calculated for the training set as well as for a randomly chosen set of test data (termination set). The ANN loses it ability to generalize when the MSE for the termination set (TS) increases. In this example, the MSE for the termination data increases after 75 epochs although the ANN is still able to improve the MSE of the TS for the training data.

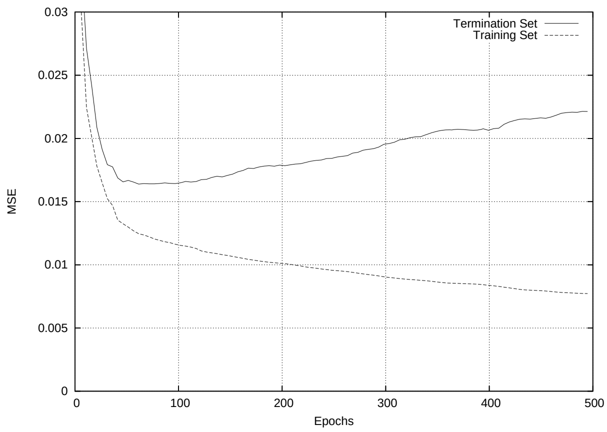


Fig. 4. MSE of training and termination set

In this study, the MSE of the ANN for the termination set is used as the termination criteria. If the MSE for the TS does not improve for at least 300 epochs, the training phase is stopped and the ANN of the epoch where the MSE was minimal is used and applied to the VS.

4.3.3 Configuration Process

We present experiments to find a high-quality ANN configuration. Different configurations are compared where only one configuration parameter (e.g. the learning function) is varied and all other configuration parameters remain the same. The forecasting quality (MSE of the VS) and if required the computation time of the training phase is used to evaluate the different configurations.

Unless specified differently, the CS, VS, and TS have equal size. Each set includes 5,000 randomly chosen itineraries. We limit the set size to 5,000 because we perform at least five training runs with different data sets for each configuration and the training runs are computationally expensive. We are aware that using a low TS size reduces the forecasting quality of ANNs. However, deciding among different configurations is not affected by the low TS size since it affects different configurations in the same way. Furthermore, the final experiments use the same number of observations as the MNL model and EIS.

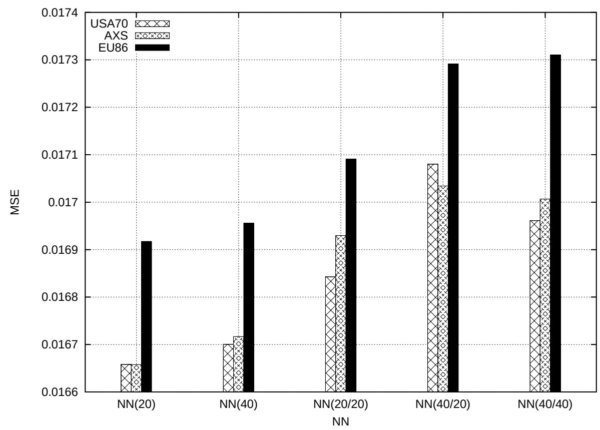


Fig. 5. MSE over different network topologies for different time preference functions

Because the topology of an ANN plays a major role in forecasting quality, basic parameters such as time preferences or learning functions are evaluated for different topologies. Five topologies are compared, consisting of different numbers of neurons and hidden layers. For example, NN(20) denotes an ANN with one layer of 20 hidden neurons, NN(40/20) denotes an ANN with two layers with 40 neurons in the first and 20 neurons in the second hidden layer.

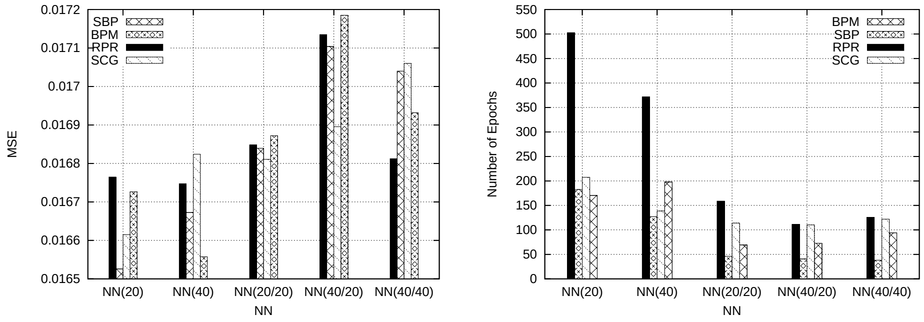
4.3.3.1 Time Preference Three different time preference functions are available to be used for the ANN (see Section 2.2.1). Our goal is to identify the time preference function that best describes the real preference of the passengers and produces the lowest MSE. Figure 5 presents results for the different time preference functions and five different network topologies. These results represent averages of training runs using the following different learning functions with standard parameters (Zell et al., 1995; Arbib, 2002):

- Standard Backpropagation Algorithm (SBP)
- Backpropagation Algorithm with Momentum Term (BPM)
- Resilient Propagation (RPR)
- Scaled Conjugate Gradient (SCG)

The numbers show that the preference function USA70 results in the lowest MSE (except for NN(40/20)) independently of the used learning function. Thus, USA70 is used for the further experiments.

4.3.3.2 Learning Function We study how the prediction quality of ANNs depend on the type of learning function. Figure 6 presents the MSE and the number of training steps for different learning functions (see above).

Because the different learning functions result in almost the same prediction

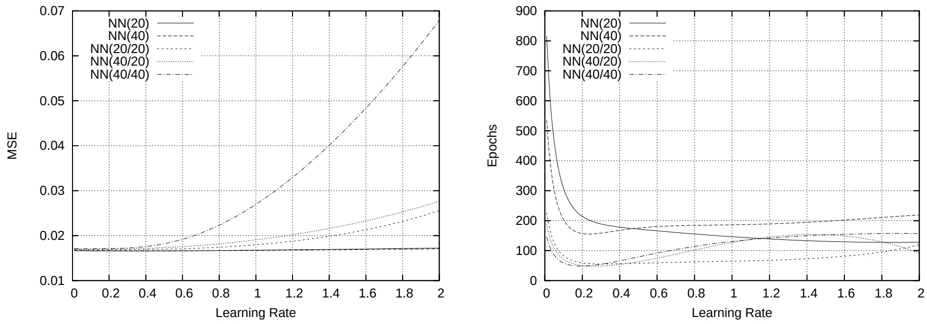


(a) MSE (b) Epochs

Fig. 6. MSE and number of training steps over different network topologies for different learning functions

quality, a standard backpropagation algorithm is used as the learning function in further experiments because it requires the lowest number of training steps.

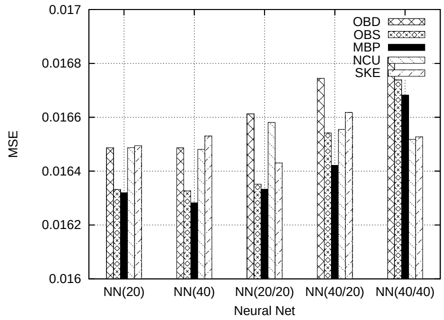
The only parameter of the standard backpropagation algorithm that has to be set by the user is the learning rate. It controls the step size of the weight adjustments. Figure 7 shows the impact of the learning rate on the MSE and average number of training steps. A learning rate of about 0.2 results in the lowest number of training steps and is used in subsequent experiments.



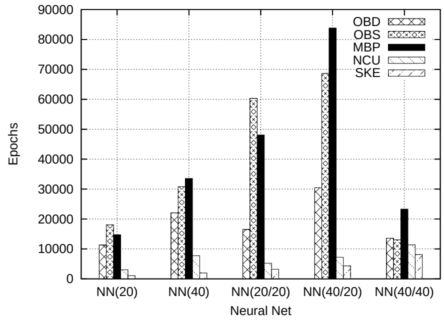
(a) MSE (b) number of training steps (epochs)

Fig. 7. MSE and number of training steps over the learning rate of a standard backpropagation algorithm

4.3.3.3 Network Topology One of the most difficult and least intuitive decisions for the design of an ANN is the choice of a proper topology. Decisions have to be made on the number of hidden layers, the number of neurons in each layer, and the connectivity between the neurons. Pruning algorithms can help in finding a high-quality topology (Lam and Stork, 2002). Pruning algorithms start with an initial topology and iteratively eliminate neurons and/or links according to specific criteria (e.g. if the weight of a link or the activation of a neuron is close to 0). In this study, we iteratively use pruning functions and train the ANN until a satisfactory topology is found. Five different pruning



(a) MSE



(b) Epochs

Fig. 8. MSE and number of training steps over different initial topologies using different pruning algorithms

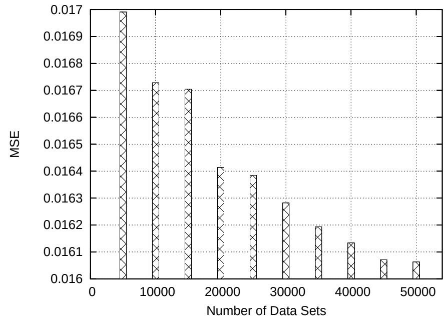
algorithms are evaluated with respect to the resulting MSE and number of training steps (see Lam and Stork (2002) for a detailed description of the pruning algorithms and further references):

(1) Magnitude Based Pruning (MBP),
(2) Optimal Brain Damage (OPD),
(3) Optimal Brain Surgeon (OBS),
(4) Non Contributing Units (NCU), and
(5) Skeletonization (SKE).

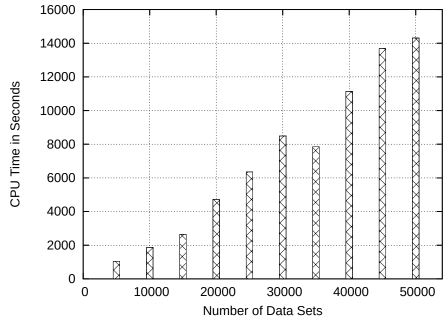
Figure 8 shows how the MSE and the number of training steps depends on the used pruning algorithm for different initial topologies.

The results show only small differences of the MSE for the different algorithms functions; the highest MSE is only 2% higher than the lowest MSE (OBD in comparison to MBP for NN(40/20)). Except for the initial topology NN(40/40), MBP always yielded the lowest MSE. However, large differences in computation time exist between the different approaches. For example, the number of training steps required by MBP is 1,800% higher than required by SKE for NN(40/20). As SKE is fastest for all topologies, and all considered pruning algorithms result in similar MSE. SKE is used as the pruning function in subsequent experiments.

4.3.3.4 Number of Data Sets In all previous experiments, the size of the CS, VS, and TS has been limited to 5,000 due to the high computation time required for calibration. In the final experiments, we study the impact of the size of the data sets on the MSE and the computation time. The same ANN is trained and evaluated with an increasing number of data sets (5 experiments each). We use the topology NN(20/20) as it yielded the lowest MSE in the previous experiments. In figure 9, the MSE and CPU-time is shown for increasing numbers of data sets.



(a) MSE



(b) CPU-time

Fig. 9. MSE and CPU-time over different sizes of the training, validation, and termination set

The results indicate that a higher number of training sets yield a better forecasting quality. On the other hand, a higher number of data sets also requires higher computation time per training step.

5 Custom Model

In this section, we present a custom model (denoted as EIS model) for the estimation of itinerary market shares. Analogously to the MNL and ANN models, we use the data to calibrate the EIS model and compare its performance to the other two presented models.

5.1 Model Description

The attraction $A_k(t)$ of an itinerary $k \in K$ at time $t \in [t_{min}, t_{max}]$ is estimated based on its attributes $\mathbf{X(t)_k} = (x_{k1}, x_{k2} \ldots, x_{ka-1}, x_{ka}(t))$ and the parameters $\boldsymbol{\beta} = (\beta_1, \beta_2 \ldots, \beta_a)$. Appropriate values for the parameters $\boldsymbol{\beta}$ are determined in the calibration phase (see Section 5.2). The attribute $x_{ka}(t)$ depends on the time-dependent departure time preference DTP. The attraction of an itinerary k is calculated by summing up the weighted attributes:

$$\begin{aligned} A_k(t) &= \boldsymbol{\beta}^T \mathbf{X(t)_k} \\ &= \beta_1 x_{k1} + \beta_2 x_{k2} + \beta_3 x_{k3} \ldots \beta_{a-1} x_{ka-1} + \beta_a x_{ka}(t). \end{aligned} \tag{1}$$

The departure time preference $DTP(t)$ of an itinerary k is weighted by the difference between the departure time dep_k of itinerary k and the preferred departure time t of a passenger. With increasing difference $|dep_k - t|$ between departure time dep_k and preferred departure time t of a passenger, $x_k(t)$ decreases according to a Gaussian function. It is calculated as

$$x_k(t) = DTP(t)\exp(-(t - dep_k)^2/\lambda),$$

where the different possibilities for $DTP(t)$ are shown in Figure 1 and λ is a free parameter that is adjusted during model calibration. The Gaussian function models that the attraction of an itinerary k decreases with increasing $|t - dep_k|$. Consequently, the attraction of itinerary k is maximal for all passengers who want to fly at time $t = dep_k$.

$S_k(t)$ is defined as the absolute, normalized attraction ($S_k(t) \in [0, 1]$) of itinerary k for any time $t \in \{t_{min}, t_{max}\}$. It is calculated as

$$S_k(t) = \frac{A_k(t)}{\max_{i \in K_m} \max_{t \in \{t_{min}, t_{max}\}}(A_i(t))},$$

where K_m is the set of itineraries in the market m where itinerary k belongs to ($k \in K_m$). Therefore, the denominator finds the maximal attraction over the time horizon of all itineraries that belong to the same market K_m as itinerary k. $S_k(t)$ represents the attraction of itinerary k independently of competing itineraries.

If there is more than one itinerary k in one market K_m a potential passenger can choose between the competing itineraries. We introduce the relative attraction $R_k(t)$ of an itinerary k as

$$R_k(t) = \frac{S_k(t)}{\max\left(\sum_{i \in K_m} S_i(t), 1\right)}.$$

Then, the total demand D_k (number of passengers) of itinerary k can be calculated as

$$D_k = \sum_t R_k(t).$$

We assume that time is discretized and $t \in \{t_{min}, t_{max}\}$.

5.2 Calibration and Validation

The goal of calibration is to find the parameters $(\beta_1, \beta_2 \ldots, \beta_a)$ and λ that yield the highest forecasting quality (minimum MSE). For this purpose we used threshold accepting (Dueck and Scheuer, 1990). Threshold accepting is an iterative optimization procedure where a new solution in the neighborhood of the current solution is selected if either solution quality increases, or the decrease in solution quality is below a given threshold. During an optimization

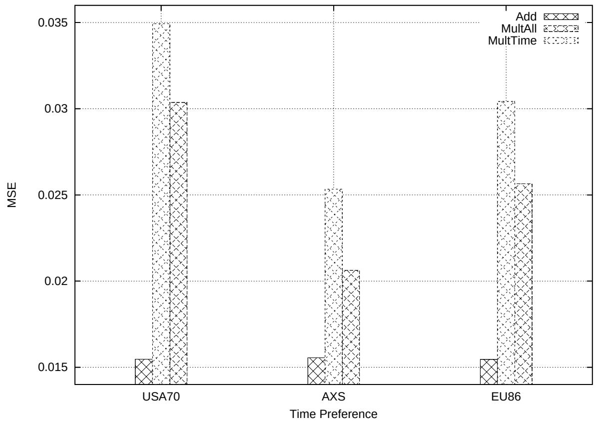


Fig. 10. MSE over different time preference functions for different EIS models

run, the threshold is reduced to zero. A solution is encoded as a vector of parameters (consisting of $(\beta_1, \beta_2 \ldots, \beta_a)$ and λ). The quality of a solution is the MSE of the calibration set. Thus, a higher MSE implicates a lower solution quality.

All variables from Table 1 are used for the EIS except CLO and TRC. These variables are excluded because their effects are implicitly included in the model structure in which the attraction is calculated for each point in time taking into account the competing itinerary's attractions. As already mentioned, λ is included as an additional variable.

Equation (1) calculates the attraction of an itinerary as the sum of its weighted attributes. We tested two other, alternative, model formulations. In the first alternative model formulation (MultAll), addition was replaced by multiplication. In the second alternative model formulation (MultTime), the attraction of an itinerary k is calculated as

$$A_k(t) = (\beta_1 x_{k1} + \ldots + \beta_{a-1} x_{ka-1})\beta_a x_{ka}.$$

Figure 10 shows the resulting MSE over different time preference functions for the three different model formulations. We performed 10 experiments with a different set of 40,000 randomly chosen itineraries as CS and VS for each alternative.

The results indicate that the additive model (denoted as Add) described in (1) yields the best results as for all three time preference functions the MSE is lower than for the alternative models. The MSE is lowest when using EU86 as the time function.

6 Results

Because all three models studied in this paper have the same objective and require the same input data, a comparison between them is straightforward. In the previous sections, we determined the best setting for each of the three different models. This section compares the resulting models in identical experimental setups. For each model, two experimental setups with different numbers of observations in the calibration set (CS) and in the validation set (VS) were conducted:

- Setup 1: The CS contains 50,000 randomly chosen itineraries. For validation, we performed 10 independent runs with 50,000 randomly chosen itineraries. Each experiment is repeated 10 times with randomly chosen data sets.
- Setup 2: The CS contains 400,000 and the VS 200,000 randomly chosen itineraries. We have been not able to use all available itineraries since the training of the ANN model requires a termination set which contains 200,000 itineraries. We performed two experiments with non-overlapping itineraries in the CS.

6.1 MNL

Table 2 presents the results (averaged $\bar{\rho^2}$ and MSE) for both experimental setups using the MNL model from Section 3. Furthermore, we show the parameter estimates $\hat{\boldsymbol{\beta}}$ for the input variables. The time preference function USA70 is used for calibration.

The results of both experimental setups show a high log-likelihood ratio index $\bar{\rho^2}$, indicating a valid model structure and good fit. The differences in the parameters estimates $\hat{\boldsymbol{\beta}}$ between both setups are small and negligible. All variables are significant on the 0.999 level in all experiments. Exceptions are one experiment of setup 1, where QUA is significant on the 0.90 level, three experiments of setup 1 where STY is on the 0.750 level, and two other experiments where STY became insignificant.

The estimates of the different variables can be interpreted with respect to their impact on an itinerary's attraction, resp. market share. In particular, the estimates for the travel time ratio (TTR) indicate a positive impact of a shorter travel time on the attraction of an itinerary. A lower travel time (representing an increase of the variable) results in an increase of the attraction of an itinerary. The same effect can be observed for the itinerary type (TYP). Passengers prefer direct flights and avoid connection flights due to the increased travel time, the inconvenience of switching planes, or higher probability of delays and lost baggage. This has also been observed by Coldren et al.

variable	$\hat{\boldsymbol{\beta}}$ (setup 1)	$\hat{\boldsymbol{\beta}}$ (setup 2)
TTR	0.337	0.355
TYP	1.996	1.970
STY	-0.009	-0.044
DTP	0.125	0.131
QUA	0.009	0.015
PRS	0.720	0.714
TRC	0.055	0.044
CLO	0.003	0.003
$\bar{\rho}^2$	0.333	0.328
MSE	0.0165	0.0161

Table 2
Results for the MNL model

(2003). This effect is especially strong because the used data sets (see Section 2.2.1) contain only short-haul routes (itineraries between Germany and European countries). When applying the model to long-haul routes, the advantage of direct flights is expected to be lower due to the reduced perceived disadvantage of connection flights in comparison to direct flights. A high positive impact on attraction can also be observed for an airlines presence in a market (PRS). This reflects the strong position of national air carriers on routes to or from their home countries. Usually, a carrier with a high presence in a market can offer more flights and get more acknowledgment from potential passengers. This was also observed by Teodorovic and Krcmar-Nozic (1989). Furthermore, the results for DTP indicate that passengers prefer departure times following the time preference function USA70. The impact of an airlines quality is low (QUA). This can be explained due to the low differences of service qualities between airlines offering air service in Europe. In addition, flights are short and quality is only an important factor for long-haul flights. Finally, all variables representing the competition in a market (STY, TRC, and CLO) have a low impact on the attraction of an itinerary.

6.2 ANN

For the ANN, setup 1 resulted in an MSE of 0.0160 and setup 2 in an MSE of 0.0157. For these experiments, ANNs were trained using the standard backpropagation algorithm with 0.2 as learning rate and pruned using skeletonization as the pruning algorithm. The initial topologies had two hidden layers with 20 neurons each (NN(20,20)) and the used time preference function was

USA70. The structure of an ANN does not allow a meaningful analysis or interpretation of the importance of the input variables.

6.3 EIS

Table 3 presents the results (average MSE) for both experimental setups using the EIS model from Section 5. Inside the EIS model, attraction values are normalized and relative attractions are used. Therefore, the absolute parameter estimates $\hat{\boldsymbol{\beta}}$ that are derived for different setups can not be compared directly. To be able to directly compare the parameter estimates for different setups (setup 1 and setup 2), we also show normalized parameter estimates $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} / \max(\hat{\boldsymbol{\beta}})$.

variable	$\hat{\boldsymbol{\beta}}$ (setup 1)	$\hat{\boldsymbol{\beta}}$ (setup 2)	$\tilde{\boldsymbol{\beta}}$ (setup 1)	$\tilde{\boldsymbol{\beta}}$ (setup 2)
TTR	0.392	0.583	0.253	0.248
TYP	1.550	2.344	1.0	1.0
STY	1.218	1.840	0.785	0.784
DTP	0.827	1.354	0.533	0.577
QUA	0.062	0.100	0.039	0.042
PRS	0.502	0.728	0.324	0.310
MSE	0.0156	0.0155		

Table 3
Final results for EIS

Both experimental setups yield similar results for the normalized parameter estimates $\tilde{\boldsymbol{\beta}}$, indicating a high stability and confidence of the model. t-tests on the parameters show significance on the 0.999 level for all variables and experiments, except QUA which is significant on the 0.950 level for one experiment of setup 1. As for the MNL, the estimates can be interpreted as the individual impact of an independent variable on an itinerary's attraction. In general, the EIS model results in estimates which are similar to those of the MNL. For example, the type of itinerary (TYP) has a high impact, whereas the quality of an airline (QUA) only has a minor effect on an itineraries' attraction. In comparison to the MNL model, the type of the shortest itinerary in the market (STY) and the departure time preference (DTP) has a stronger impact. This is due to the different attraction calculation inside the EIS model where the variables TRC and CLO are not used and only considered implicitly in the time-dependent variable $x_k(t)$. As discussed in Section 5.1, λ was also a subject of estimation in the final experiments. Setup 1 resulted in $\lambda = 5.937$ and setup 2 resulted in $\beta = 5.951$.

Table 4 compares the mean and standard deviation (in brackets) of the MSE of the two different setups for the different models

	MNL	ANN	EIS
setup 1	0.0165 (1.6×10^{-4})	0.0160 (0.3×10^{-4})	0.0155 (0.2×10^{-4})
setup 2	0.0161 (5.4×10^{-4})	0.0157 (0.2×10^{-4})	0.0156 (2.4×10^{-4})

Table 4
Mean MSE for different forecasting models (standard deviations are in brackets)

The results indicate that the EIS model outperforms the two other models. The MNL model results in the lowest prediction quality.

To confirm these observations, we perform an unpaired t-test for setup 1. The null hypothesis H_0 is that the observed differences in the forecasting quality (EIS outperforms MNL and ANN) are random. H_α says that the differences are a result of the model specification. The critical t-value for $p = 0.999$ is 3.6105. The results shown in Table 5 for the three models show that the t-values always exceed the critical t-value. Thus, H_0 can be rejected on the 99.9%-level.

models	t-value
MNL vs. ANN	10.1881
MNL vs. EIS	18.9338
ANN vs. EIS	42.6303

Table 5
t-values

7 Summary and Conclusions

In this paper, three different forecasting methods for air travel itinerary market share estimation were formulated, calibrated, and tested using historical booking data: a multinomial logit (MNL) model, an artificial neural network (ANN), and a custom model (EIS) developed by the authors. In calibration, each model is calibrated such that it best reproduces historical booking data.

All models yielded stable prediction results. With respect to the minimum squared error which is used to evaluate the prediction quality of the models, the EIS model outperformed the two other models. The MNL model showed lowest prediction quality.

The presented results recommend an increased use of the presented EIS model for passenger prediction as this model resulted in the highest prediction accuracy. Furthermore, in contrast to ANN models, in the EIS model the input variables individual impact on the output variable can be extracted. Thus, airlines can focus on these individual variables for the development of appropriate strategies to increase their itineraries attraction.

References

Alamdari, F. E., Black, I. G., 1992. Passengers' choice of airline under competition: The use of the logit model. Transport Reviews 12 (2), 153–170.

Arbib, M. A., 2002. The elements of brain theory and neural networks. In: Arbib, M. A. (Ed.), The Handbook of Brain Theory and Neural Networks, 2nd Edition. The MIT Press, Cambridge, London, pp. 1–23.

Ashford, N., Benchemam, M., 1987. Passengers' choice of airport: An application of the multinomial logit model. Transportation Research Record 1147, 1–5.

Ben-Akiva, M., Bierlaire, M., 1999. Discrete choice methods and their application to short term travel decisions. In: Hall, R. (Ed.), Handbook of Transportation Science. Vol. 23.

Ben-Akiva, M., Lerman, S. R., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press, Cambridge, Massachusetts.

Biermann, T., 1986. Die Bedeutung der "sechsten Freiheit" für den Wettbewerb im Luftverkehr. Vol. 46 of Buchreihe des Instituts für Verkehrswissenschaft an der Universität zu Köln. Köln.

Coldren, G. M., Koppelman, F. S., 2005. Modeling the competition among air-travel itinerary shares: GEV model development. Transportation Research Part A: Policy and Practice 39, 345–365.

Coldren, G. M., Koppelman, F. S., Kasturirangan, K., Mukherjee, A., 2003. Modeling aggregate air-travel itinerary shares: Logit model development at a major US airline. Journal of Air Transport Management 9, 361–369.

Dueck, Scheuer, 1990. Threshold accepting: A general purpose optimization algorithm appearing superior to simulated annealing. Journal of Computational Physics 90 (1), 161–175.

Hruschka, H., 1993. Determining market response function by neural network modeling: A comparison to econometric techiques. European Journal of Operational Research 66, 28.

Hsu, C.-I., Wen, Y.-H., 2003. Determining flight frequencies on an airline network with demand-supply interactions. Transportation Research Part E: Logistics and Transportation Review 39 (6), 417–442.

Kanafani, A., 1983. Transportation demand analysis. McGraw-Hill Book Company, New York.

Lam, C. P., Stork, D. G., 2002. Learning network topology. In: Arbib, M. A.

(Ed.), The Handbook of Brain Theory and Neural Networks, 2nd Edition. The MIT Press, Cambridge, London, pp. 628–631.

McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (Ed.), Frontiers in Economics. Academic Press, New York, London, pp. 105–142.

Nam, K., Yi, J., Prybutok, V. R., 1997. Predicting airline passenger volume. The Journal of Business Forecasting (Spring), 14–16.

O'Connor, W. E., 1982. An introduction to airline economics, 3rd Edition. Praeger, New York.

Skytrax, 2006. World airline star ranking. `http://www.airlinequality.com/StarRanking/ranking.htm`, May 22, 2006.

Teodorovic, D., Krcmar-Nozic, E., 1989. Multicriteria model to determine flight frequencies on an airline network under competitive conditions. Transportation Science 23 (1), 14–25.

Train, K. E., 2003. Discrete Choice Methods with Simulation. Cambridge University Press.

Urban, D., 1993. Logit-Analyse - Statistische Verfahren zur Analyse von Modellen mit qualitativen Response-Variablen. G. Fischer, Stuttgart, Jena, New York.

Weatherford, L. R., Gentry, T. W., Wilamowski, B., 2003. Neural network forecasting for airlines: A comparative analysis. Journal of Revenue and Pricing Management 1 (4), 319–331.

Zell, A., Mamier, G., Vogt, M., Mache, N., Hübner, R., Döring, S., Herrmann, K.-U., Soyez, T., Schmalzl, M., Sommer, T., Hatzigeorgiou, A., Posselt, D., Schreiner, T., Kett, B., Clemente, G., Wieland, J., 1995. SNNS Stuttgart Neural Network Simulator user manual, version 4.1.