# An air itinerary choice model based on a mixed RP/SP dataset

Bilge Atasoy *   Michel Bierlaire *

April 26, 2012

Report TRANSP-OR 120426
Transport and Mobility Laboratory
Ecole Polytechnique Fédérale de Lausanne
transp-or.epfl.ch

## Abstract

In this paper, we present an itinerary choice model based on a mixed RP/SP dataset. The aim of the combination of the two datasets is to exploit the variability of the SP data for the estimation of the RP model parameters. As a result a price elastic demand model is obtained which will be integrated in an airline schedule planning framework. This integration will enable us to explicitly model the supply-demand interactions which is critical for airlines for superior schedule planning decisions.

**Keywords:** Discrete choice modeling, air itinerary choice, revealed preferences, stated preferences, combined revealed preferences and stated preferences data, mixed data sources

## 1 Introduction

Demand forecasting models of airlines are critical in a profitable planning of the network and schedule. In the last decade discrete choice methodology has been introduced in the

*Transport and Mobility Laboratory (TRANSP-OR), School of Architecture, Civil and Environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland, {bilge.kucuk, michel.bierlaire}@epfl.ch

context of demand analysis of airlines (Garrow, 2010). It has been shown by Coldren et al. (2003) that discrete choice modeling leads to superior forecasts compared to a widely used Quality Service Index (QSI).

In air transportation context an *itinerary* is defined as a product between an origin and destination pair which can be composed of several flight legs. Since the information on the demand is available on the itinerary level, choice models are developed for the itineraries. In the literature, random utility models have been used to model the choice of itinerary depending on various attributes. We refer to the work of Garrow (2010) for a comprehensive review of different specifications of choice behavior models for air travel demand. Coldren et al. (2003) propose logit models and Coldren and Koppelman (2005) extend the previous work with the introduction of GEV and nested logit models. Gramming et al. (2005) propose a probit model where there is a large set of alternatives in the context of non-IIA problems. Koppelman et al. (2008) model the time of day preferences under a logit setting in order to analyze the effect of schedule delay. Carrier (2008) and Wen and Lai (2010) propose some advanced demand modeling in which customer segmentation is modeled as a latent class.

In this study we develop an itinerary choice model based on a real dataset. The dataset is a mixed revealed preferences (RP) and stated preferences (SP) dataset. The RP data is a booking data from a major European airline and the SP data is based on an Internet survey in US. The contribution of this study is the price elasticity of the resulting demand model which is lacking in the models based on booking data. This demand model is aimed to be integrated with a schedule planning model. This integration provides simultaneous decisions on the schedule plan and revenue management.

The rest of the paper is organized as follows. Section 2 presents the itinerary choice model we develop. Section 3 presents the mixed RP/SP data used for the estimation. The methodology for the joint estimation of RP and SP models is presented in section 4. We provide the estimation results in section 5 together with the indicators of demand including willingness to pay and elasticities. Finally, we conclude in section 6.

# 2 Itinerary choice model

We develop an itinerary choice model for the choice of alternative itineraries in the same market segments. The market segments , $s \in S^h$, are defined by the origin and destination (OD) pairs and they are differentiated for each cabin class $h$. Considered classes are economy and business classes and therefore we have two segments for each OD pair. The choice situation is defined for each segment $s$ with a choice set of all the alternative itineraries in the segment represented by $I_s$. The index $i$ for each alternative itinerary in segment $I_s$ carries the information of the cabin class of the itinerary due to the definition of the segments. Therefore we do not use the index $h$ for the itineraries. As an example, consider a market segment of economy passengers between Geneva and Washington. The alternatives for this segment includes all the available economy itineraries which can be non-stop or connecting itineraries with different departure times. Finally, in order to better represent the reality, we include *no-revenue options* ($I_s^{'} \subset I_s$), which represent the itineraries offered by competitive airlines.

The utility of each alternative itinerary $i$, including the no-revenue options, is represented by $V_i$ and the specification is provided in Table 1. The alternative specific constants, $ASC_i$, are included for each itinerary in each segment except one of them which is normalized to 0 for identification purposes. Other parameters are represented by $\beta$ for each of the explanatory variables. Since we have different models for economy and business classes all the parameters and variables are specified accordingly. Superscripts $E$ and $B$ indicate the economy and business classes respectively. Moreover, the superscripts $NS$ and $S$ indicate the non-stop and one-stop itineraries respectively. The explanatory variables are described as follows:

- $price_i$ is the price of itinerary $i$ in €, which is normalized by 100 for scaling purposes,
- $time_i$ is the elapsed time for itinerary $i$ in hours,
- $non\text{-}stop_i$ is a dummy variable which is 1 if itinerary $i$ is a non-stop itinerary, 0 otherwise,
- $stop_i$ is a dummy variable which is 1 if itinerary $i$ is a one-stop itinerary, 0 otherwise,
- $economy_i$ is a dummy variable which is 1 if itinerary $i$ is an economy itinerary, 0 otherwise,
- $business_i$ is a dummy variable which is 1 if itinerary $i$ is a business itinerary, 0 otherwise,
- $morning_i$ is a dummy variable which is 1 if itinerary $i$ is a morning itinerary departing between 07:00-11:00, 0 otherwise. The time slot is inspired by the studies in literature that show that the individuals have higher utility for the departures in this slot (Garrow, 2010).

As seen in Table 1 all the parameters are interacted with the *economy* and *business* dummies in order to be able to have two different models for the two classes. In addition to the interaction with the cabin class, the time and price variables are interacted with the number of stops, i.e. the dummies of *non-stop* and *stop* since there are strong correlations between the number of stops and both the time and price of the itinerary. Furthermore, the price variable is included as a log formulation since it improved the model significantly. The idea behind is that, the effect of the increase in price is not linear for different levels of the price.

Table 1: Specification table of the utilities

| | Parameters | Explanatory variables |
|---|---|---|
| constants | $\mathrm{ASC}_i^E$ | $1 \times \mathrm{economy}_i$ |
| | $\mathrm{ASC}_i^B$ | $1 \times \mathrm{business}_i$ |
| price | $\beta_{\mathrm{price}}^{E,NS}$ | $\ln(\mathrm{price}_i/100) \times \text{non-stop}_i \times \mathrm{economy}_i$ |
| | $\beta_{\mathrm{price}}^{B,NS}$ | $\ln(\mathrm{price}_i/100) \times \text{non-stop}_i \times \mathrm{business}_i$ |
| | $\beta_{\mathrm{price}}^{E,S}$ | $\ln(\mathrm{price}_i/100) \times \mathrm{stop}_i \times \mathrm{economy}_i$ |
| | $\beta_{\mathrm{price}}^{B,S}$ | $\ln(\mathrm{price}_i/100) \times \mathrm{stop}_i \times \mathrm{business}_i$ |
| time | $\beta_{\mathrm{time}}^{E,NS}$ | $\mathrm{time}_i \times \text{non-stop}_i \times \mathrm{economy}_i$ |
| | $\beta_{\mathrm{time}}^{B,NS}$ | $\mathrm{time}_i \times \text{non-stop}_i \times \mathrm{business}_i$ |
| | $\beta_{\mathrm{time}}^{E,S}$ | $\mathrm{time}_i \times \mathrm{stop}_i \times \mathrm{economy}_i$ |
| | $\beta_{\mathrm{time}}^{B,S}$ | $\mathrm{time}_i \times \mathrm{stop}_i \times \mathrm{business}_i$ |
| time-of-day | $\beta_{\mathrm{morning}}^{E}$ | $\mathrm{morning}_i \times \mathrm{economy}_i$ |
| | $\beta_{\mathrm{morning}}^{B}$ | $\mathrm{morning}_i \times \mathrm{business}_i$ |

The choice model is formulated as a logit model. It gives the choice probability for each itinerary $i$ in segment $s$ as represented by equation 1.

$$P^s(i) = \frac{\exp(V_i)}{\sum_{j \in I_s} \exp(V_j)} \qquad \forall h \in H, s \in S^h, i \in I_s \tag{1}$$

# 3 Data

For the estimation of the demand model we use an RP data provided in the context of ROADEF Challenge 2009[1]. This is a booking data from a major European airline which provides the set of airports, flights, aircraft and itineraries. The information provided for the itineraries includes the corresponding flight legs therefore we can deduce the information on the departure and arrival time of itinerary, the trip length and the number of stops. Additionally, we have information on the demand and average price (€) for each cabin class. This RP data does not include any information concerning the competitive airlines. Therefore the no-revenue options are not considered in the estimation process. However for applying the model we assume that these itineraries have the same type of

[1] `http://challenge.roadef.org/2009/en`

utility functions as presented in Table 1 and their attributes are assigned according to the other available itineraries in the market offered by competitive airlines.

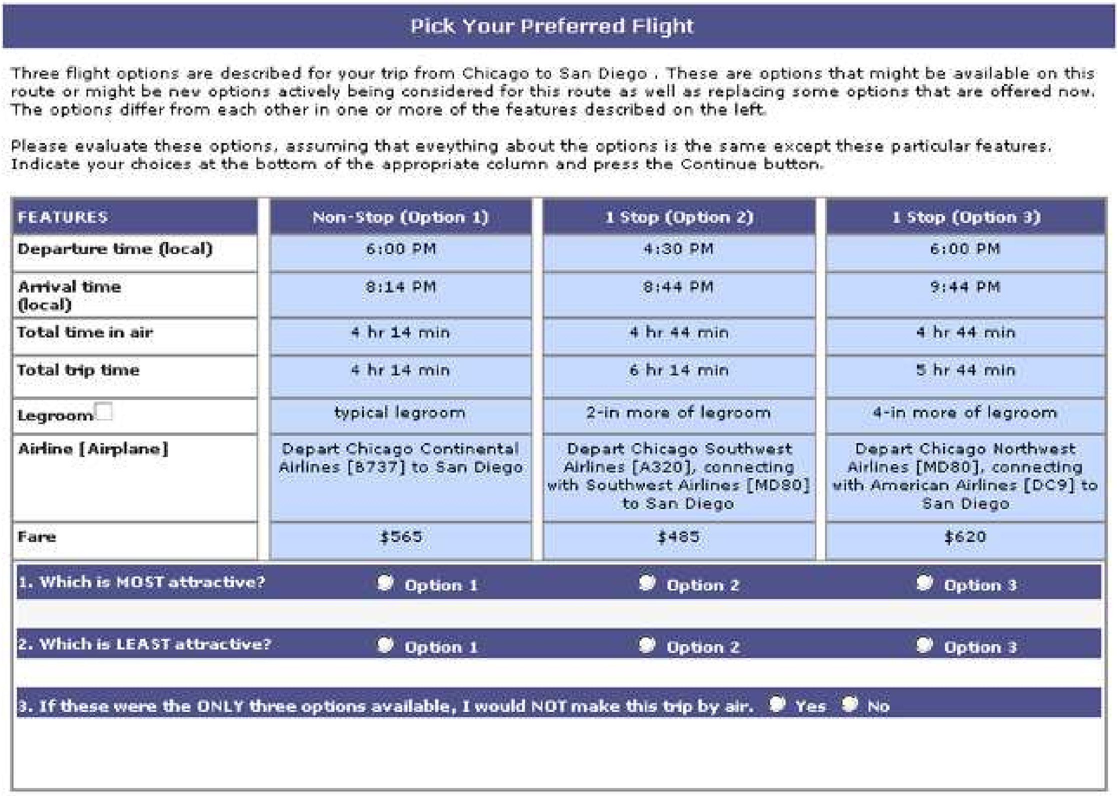
As it is common with RP data, the lack of variability in some attributes precludes a statistically significant estimation of key parameters of the choice models. Many models in literature, which are based on airline booking data, have insignificant price parameters (Garrow, 2010). Therefore, in this study, the RP data is combined with SP data, where the variability is obtained by design. In literature mixed RP/SP datasets are utilized for the estimation of discrete choice models with several purposes. Ben-Akiva and Morikawa (1990) introduce the methodology for the combination of different data sets in order to exploit the advantages of one to overcome the shortcomings of the other. Louviere et al. (1999) is a review for the usage of joint preferences datasets.

The SP data, which is used to overcome the inelastic behavior of RP data, is based on an Internet choice survey collected in 2004 in the US. Let us note that, the combined dataset therefore contains both European and US data. The Internet survey was organized to understand the sensitivity of air passengers to the attributes of an airline itinerary such as fare, travel time, number of stops, legroom, and aircraft. The respondents were presented hypothetical choice situations and offered three alternatives as seen in Figure 1. The first is a non-stop itinerary, the second one is a one-stop itinerary with the same airline and the third is connecting with a different airline. By design, the data has enough variability in terms of price and other variables. The SP sample has 3609 observations.

# 4 The simultaneous estimation of RP and SP models

As mentioned previously the RP model presented in section 2 is simultaneously estimated with the SP model in order to take the advantage of its elasticity. The SP model is also a logit model. The choice set consists of three alternatives. The first one is a nonstop itinerary. The second alternative is a one-stop itinerary both flights being operated by the same airline. The third alternative is also a one-stop itinerary where the connection is provided by another airline. The utilities for these alternatives are provided by the equations 2, 3, and 4 respectively.

Since the models for RP and SP datasets are estimated simultaneously, we need to define a scale variable, $scale_{SP}$. The the scale of the RP data is fixed to 1 and $scale_{SP}$ is to be estimated in order to capture the differences in the covariance structure of the error terms of the two models.



**Pick Your Preferred Flight**

Three flight options are described for your trip from Chicago to San Diego . These are options that might be available on this route or might be nev options actively being considered for this route as well as replacing some options that are offered now. The options differ from each other in one or more of the features described on the left.

Please evaluate these options, assuming that eveything about the options is the same except these particular features. Indicate your choices at the bottom of the appropriate column and press the Continue button.

| FEATURES | Non-Stop (Option 1) | 1 Stop (Option 2) | 1 Stop (Option 3) |
|---|---|---|---|
| Departure time (local) | 6:00 PM | 4:30 PM | 6:00 PM |
| Arrival time (local) | 8:14 PM | 8:44 PM | 9:44 PM |
| Total time in air | 4 hr 14 min | 4 hr 44 min | 4 hr 44 min |
| Total trip time | 4 hr 14 min | 6 hr 14 min | 5 hr 44 min |
| Legroom | typical legroom | 2-in more of legroom | 4-in more of legroom |
| Airline [Airplane] | Depart Chicago Continental Airlines [B737] to San Diego | Depart Chicago Southwest Airlines [A320], connecting with Southwest Airlines [MD80] to San Diego | Depart Chicago Northwest Airlines [MD80], connecting with American Airlines [DC9] to San Diego |
| Fare | $565 | $485 | $620 |

1. Which is MOST attractive? Option 1 Option 2 Option 3

2. Which is LEAST attractive? Option 1 Option 2 Option 3

3. If these were the ONLY three options available, I would NOT make this trip by air. Yes No

Figure 1: An example page for the SP survey

Similar to the RP model, the parameters are specified as economy and business. The parameters of the price variables for each of the alternatives ($\beta^{E,NS}_{\text{price}}$, $\beta^{B,NS}_{\text{price}}$, $\beta^{E,S}_{\text{price}}$, $\beta^{B,S}_{\text{price}}$) are constrained to be the same as the price parameters of the RP model presented in section 2. Similarly the parameters of the time variables ($\beta^{E,NS}_{\text{time}}$, $\beta^{B,NS}_{\text{time}}$, $\beta^{E,S}_{\text{time}}$, $\beta^{B,S}_{\text{time}}$) and the parameters of the morning variables ($\beta^{E}_{\text{morning}}$, $\beta^{B}_{\text{morning}}$) are also designed to be the same as the parameters of the RP model.

In the SP model, there are additional explanatory variables since it is based on a rich data set. For business passengers we have the information whether they pay their ticket or their company pay for that. Therefore there is an additional dummy variable, *business/others-pay*, which is 1 if the business passenger's ticket is not paid by himself. There are other explanatory variables which are represented by vector $v$. These variables include the legroom provided in the airplane, the delay of the flight in case of late or early arrival and the variable representing whether the passenger is a frequent flyer or not.

$$
\begin{aligned}
V_1 = \text{scale}_{SP} \times (&\beta^{E,NS}_{\text{price}} \times \ln(\text{price}_1/100) \times \text{economy} + \beta^{B,NS}_{\text{price}} \times \ln(\text{price}_1/100) \times \text{business} \\
&+ \beta^{B-OP}_{\text{price}} \times \ln(\text{price}_1/100) \times \text{business/others-pay} \\
&+ \beta^{E,NS}_{\text{time}} \times \text{time}_1 \times \text{economy} + \beta^{B,NS}_{\text{time}} \times \text{time}_1 \times \text{business} \\
&+ \beta^{E}_{\text{morning}} \times \text{morning}_1 \times \text{economy} \\
&+ \beta^{B}_{\text{morning}} \times \text{morning}_1 \times \text{business} \\
&+ \sum_i \beta^E_i \times v^i_1 \times \text{economy} + \beta^B_i \times v^i_1 \times \text{business})
\end{aligned} \tag{2}
$$

$$
\begin{aligned}
V_2 = \text{scale}_{SP} \times (&\text{ASC}^E_2 \times \text{economy} + \text{ASC}^B_2 \times \text{business} \\
&+ \beta^{E,S}_{\text{price}} \times \ln(\text{price}_2/100) \times \text{economy} + \beta^{B,S}_{\text{price}} \times \ln(\text{price}_2/100) \times \text{business} \\
&+ \beta^{B-OP}_{\text{price}} \times \ln(\text{price}_2/100) \times \text{business/others-pay} \\
&+ \beta^{E,S}_{\text{time}} \times \text{time}_2 \times \text{economy} + \beta^{B,S}_{\text{time}} \times \text{time}_2 \times \text{business} \\
&+ \beta^{E}_{\text{morning}} \times \text{morning}_2 \times \text{economy} \\
&+ \beta^{B}_{\text{morning}} \times \text{morning}_2 \times \text{business} \\
&+ \sum_i \beta^E_i \times v^i_2 \times \text{economy} + \beta^B_i \times v^i_2 \times \text{business})
\end{aligned} \tag{3}
$$

$$
\begin{aligned}
V_3 = \text{scale}_{SP} \times (&\text{ASC}^E_3 \times \text{economy} + \text{ASC}^B_3 \times \text{business} \\
&+ \beta^{E,S}_{\text{price}} \times \ln(\text{price}_3/100) \times \text{economy} + \beta^{B,S}_{\text{price}} \times \ln(\text{price}_3/100) \times \text{business} \\
&+ \beta^{B-OP}_{\text{price}} \times \ln(\text{price}_3/100) \times \text{business/others-pay} \\
&+ \beta^{E,S}_{\text{time}} \times \text{time}_3 \times \text{economy} + \beta^{B,S}_{\text{time}} \times \text{time}_3 \times \text{business} \\
&+ \beta^{E}_{\text{morning}} \times \text{morning}_3 \times \text{economy} \\
&+ \beta^{B}_{\text{morning}} \times \text{morning}_3 \times \text{business} \\
&+ \sum_i \beta^E_i \times v^i_3 \times \text{economy} + \beta^B_i \times v^i_3 \times \text{business})
\end{aligned} \tag{4}
$$

# 5 Estimation results

From the RP data 3 OD pairs are selected to be combined with the SP data. There are in total 30 alternative itineraries serving 904 passengers between these 3 OD pairs. These OD pairs are the ones with the most variability in the attributes. In Appendix B we provide results with 24 OD pairs from the RP data where the lack of price elasticity can be observed.

The characteristics of the alternatives for the selected 3 OD pairs can be seen in Table 5. When there is a business itinerary it is in fact the same product with the subsequent economy itinerary. For example, alternative 7 and 8 of the first OD pair are the same

Table 2: The attributes of the alternative itineraries for the RP data

| | alt. | stops | class | price (€) | time(min) | morning | Actual demand |
|---|---|---|---|---|---|---|---|
| **OD1** | 1 | one-stop | E | 563.8 | 260 | 1 | 3 |
| | 2 | one-stop | E | 312.5 | 260 | 1 | 6 |
| | 3 | one-stop | E | 262.5 | 360 | 0 | 27 |
| | 4 | non-stop | E | 175 | 70 | 0 | 49 |
| | 5 | non-stop | E | 175 | 70 | 0 | 56 |
| | 6 | non-stop | E | 175 | 70 | 1 | 38 |
| | 7 | non-stop | B | 409.5 | 70 | 1 | 9 |
| | 8 | non-stop | E | 175 | 70 | 1 | 29 |
| | 9 | non-stop | B | 409.5 | 70 | 0 | 16 |
| | 10 | non-stop | E | 175 | 70 | 0 | 26 |
| | 11 | non-stop | B | 409.5 | 70 | 0 | 2 |
| | 12 | non-stop | E | 175 | 70 | 0 | 28 |
| **OD2** | 1 | one-stop | E | 250 | 175 | 1 | 17 |
| | 2 | non-stop | E | 150 | 60 | 1 | 29 |
| | 3 | non-stop | E | 150 | 60 | 0 | 2 |
| | 4 | non-stop | E | 150 | 60 | 0 | 19 |
| | 5 | one-stop | B | 953 | 235 | 1 | 1 |
| | 6 | one-stop | E | 601.2 | 235 | 1 | 2 |
| | 7 | one-stop | B | 701.8 | 235 | 1 | 2 |
| | 8 | one-stop | E | 350 | 235 | 1 | 3 |
| **OD3** | 1 | one-stop | B | 655.5 | 265 | 1 | 4 |
| | 2 | one-stop | E | 387.5 | 265 | 1 | 6 |
| | 3 | non-stop | E | 237.5 | 95 | 1 | 59 |
| | 4 | non-stop | E | 237.5 | 95 | 0 | 125 |
| | 5 | one-stop | E | 609.8 | 230 | 0 | 3 |
| | 6 | one-stop | E | 325 | 230 | 0 | 6 |
| | 7 | non-stop | E | 237.5 | 95 | 0 | 73 |
| | 8 | non-stop | E | 237.5 | 95 | 0 | 84 |
| | 9 | non-stop | E | 237.5 | 95 | 0 | 73 |
| | 10 | non-stop | E | 237.5 | 95 | 0 | 107 |

product with different classes. In this section we provide results for the RP model since the focus of the study is to obtain an appropriate model for the RP data in order to be used in the framework of schedule planning models.

The estimation of the parameters for the joint RP/SP model is done with a maximum likelihood estimation using the software BIOGEME (Bierlaire and Fetiarison, 2009). In Table 3 we present the estimated parameters. As mentioned previously our focus is the RP results therefore we present the RP model parameters which are constrained to be common with the SP model. In addition to the common parameters we also present the scale parameter introduced in the SP model. The main observations can be listed as follows:

- The cost and time parameters have negative signs as expected since the increase in the price or the time of an itinerary decreases its utility.

- Economy demand is more sensitive to price and less sensitive to time compared to business demand as expected (Belobaba et al., 2009).

- For non-stop itineraries time and cost parameters are higher in absolute value compared to one-stop itineraries. Therefore, passengers on connecting itineraries are less affected by 1 € increase in price or 1 minute increase in travel time compared to non-stop itineraries. The reason is that, in our RP data the connecting itineraries are more expensive and by nature have longer travel time. Therefore we need to check the indicators of willingness to pay and elasticities to analyze these effects appropriately.

- Departure time of the day parameter, $\beta_{morning}$, is higher for business demand compared to the economy demand, which means that business passengers have a higher tendency to chose morning itineraries.

- Scale parameter for the SP model is significant and has a value of 4.32 which indeed confirms that the variability of the SP data is higher than the RP data.

- In the SP model there is an additional price parameter, $\beta_{\text{price}}^{B-OP}$, for the individuals who do not pay their ticket. This parameter has a positive sign which says that people have higher utility when their tickets are paid by their companies as expected.

- All the parameters are significance with a 90% confidence level except the time-of-day parameter for economy class.

In order to see the added value of the combination of the two datasets, in Table 4 we present the estimated values of the same parameters when using only the RP data. It is seen that the parameters are not significant which prevents us from drawing conclusions. Even the sign of the parameters are inconsistent with reality. Therefore, the model based on the RP data cannot be used for forecasting future market shares of the itineraries.

In Appendix A in Table 7 we present the results estimated with the SP data. In order to have a comparison, in Table 8 we provide the scaled values for the joint estimation results. It is seen that the results of the joint dataset is close to that of the SP data. Therefore when we have only 3 OD pairs for the RP data, the results are mainly guided by the SP data. Especially when we look at the price parameters, SP data is dominant since RP data does not have enough variability. For the time parameters the RP data has

Table 3: Estimated parameters for the model with joint RP and SP data

| | Parameters | Estimated value | t-test |
|---|---|---|---|
| RP & SP | $\beta^{E,NS}_{\text{price}}$ | -2.23 | -3.48 |
| | $\beta^{B,NS}_{\text{price}}$ | -1.97 | -3.64 |
| | $\beta^{E,S}_{\text{price}}$ | -2.17 | -3.48 |
| | $\beta^{B,S}_{\text{price}}$ | -1.97 | -3.68 |
| | $\beta^{E,NS}_{\text{time}}$ | -0.102 | -2.85 |
| | $\beta^{B,NS}_{\text{time}}$ | -0.104 | -2.43 |
| | $\beta^{E,S}_{\text{time}}$ | -0.0762 | -2.70 |
| | $\beta^{B,S}_{\text{time}}$ | -0.0821 | -2.31 |
| | $\beta^{E}_{\text{morning}}$ | 0.0283 | 1.21* |
| | $\beta^{B}_{\text{morning}}$ | 0.0790 | 1.86 |
| SP | $\text{scale}_{SP}$ | 4.32 | 3.50 |
| | $\beta^{B-OP}_{\text{price}}$ | 0.813 | 2.91 |

(* Statistical significance < 90%)

an effect on the results. However when more RP observations are included as presented in Appendix B the results change significantly.

Table 4: Estimated parameters based on the RP data

| Parameters | Estimated value | t-test |
|---|---|---|
| $\beta^{E,NS}_{\text{price}}$ | 0.0851 | 0.08* |
| $\beta^{B,NS}_{\text{price}}$ | -0.451 | -0.60* |
| $\beta^{E,S}_{\text{price}}$ | -1.47 | -0.88* |
| $\beta^{B,S}_{\text{price}}$ | -3.19 | -1.63* |
| $\beta^{E,NS}_{\text{time}}$ | -0.0204 | -0.31* |
| $\beta^{B,NS}_{\text{time}}$ | -0.108 | -1.04* |
| $\beta^{E,S}_{\text{time}}$ | -0.0705 | -0.12* |
| $\beta^{B,S}_{\text{time}}$ | 0.969 | 1.13* |
| $\beta^{E}_{\text{morning}}$ | 0.282 | 0.34* |
| $\beta^{B}_{\text{morning}}$ | -0.700 | -0.85* |

(* Statistical significance < 90%)

Since this is a complicated model with the combination of two datasets, it is better to analyze the demand indicators such as willingness to pay and elasticities rather than the parameter estimates themselves.

## 5.1 Value of time

Value of time (VOT) is the willingness of passengers to pay for one hour of travel. For each alternative $i$ VOT is given by equation 5. Since the price is included as a log formulation in the utilities VOT formula includes the price.

$$\begin{aligned} VOT_i &= \frac{\partial V_i/\partial time_i}{\partial V_i/\partial price_i} \\ &= \frac{\beta_{time} \cdot price_i}{\beta_{price}} \end{aligned} \tag{5}$$

In Table 5 the VOT values for all the alternatives of the RP data are listed. VOT is higher for business itineraries compared to economy itineraries. As an example itinerary 7 and 8 of the first OD pair can be given. This is also observed for the itineraries 9-10 and 11-12 for the first OD pair; itineraries 5-6 and 7-8 for the second OD pair; and itineraries 1-2 for the third. When we compare the VOT for non-stop and one-stop itineraries it seems as if the passengers are ready to pay more for the one-stop itineraries compared to non-stop itineraries. However this happens due to the fact that one-stop itineraries are more expensive.

Therefore in order to see the effect of the number of stops in VOT we consider two itineraries with the same price. As an example, let's take a non-stop and a one-stop itinerary which have the same price, 600 €. When we calculate the VOT, we observe that passengers are ready to pay 28 € for an hour reduction in the travel time of the non-stop alternative. For the one-stop itinerary this value is 21 € which is lower as expected.

## 5.2 Price and time elasticities of demand

Elasticities of demand give the sensitivity of passengers to the corresponding case. In this study we are interested in the price and time elasticities. They are given by the following equations:

$$\begin{aligned} E^{P_i}_{price_i} &= \frac{\partial P_i}{\partial price_i} \cdot \frac{price_i}{P_i} \\ E^{P_i}_{time_i} &= \frac{\partial P_i}{\partial time_i} \cdot \frac{time_i}{P_i} \end{aligned}$$

Belobaba et al. (2009) provide a range of airline O-D market price elasticities from -0.8 to -2.0. For business demand the average is given as -0.8 which means that if there is a 1% increase in cost, business demand will decrease by 0.8%. For economy demand this value is provided as -1.6. For time elasticity they mention that business demand has a time elasticity $< -1.0$ and for economy demand it is $> -1.0$ meaning that business demand is more elastic to time compared to economy demand.

The price and time elasticities are presented in Table 5 for the alternatives of the RP data. Price elasticity is higher for economy alternatives compared to business itineraries.

For example, for the first OD pair, for the alternatives 7-8, 9-10, and 11-12 economy demand is more elastic to price compared to the business demand. This phenomenon is also observed for the itineraries 5-6 and 7-8 of the second OD pair and for the itineraries 1-2 of the third OD pair. Furthermore, the elasticity is higher for the one-stop itineraries compared to non-stop itineraries. This means that, in case of an increase in price, passengers have a higher tendency to reject flying with a one-stop itinerary compared to a non-stop alternative. This is in line with the studies in literature (Garrow, 2010).

Time elasticities are low compared to the literature for the RP data since it includes European itineraries and the time attribute does not differ between different itineraries. However when we look at the relative elasticities for business and economy alternatives, it is seen that business demand is more elastic to time which is consistent with the empirical studies mentioned in Belobaba et al. (2009). Similarly, the time elasticity is higher for one-stop alternatives compared to non-stop ones which says that passengers are more sensitive to changes in the time for one-stop alternatives as expected.

### 5.3 Illustration for the application of the model

The developed demand model will be integrated in a schedule planning framework for airlines. Therefore in this section we illustrate how the model will be applied.

The alternative specific constants $\text{ASC}_i$ for each itinerary $i$ are not used for applying the model. The critical parameters for the application of the model are the price, the time and the time-of-day parameters which are kept the same for RP and SP models. The no-revenue itineraries, which are described in section 2 are introduced based on average market prices for competitor airlines.

For illustration purposes, we choose an arbitrary OD pair A-B. There are two alternatives of economy itineraries which are both nonstop itineraries. We include the no-revenue itinerary A-B$'$. The values of attributes can be seen in Table 6. According to the attributes the resulting choice probability, which is referred as the *market share*, is presented in the last column. The itinerary 2 has the lowest price and is a morning itinerary. Therefore it attracts the biggest number of passengers.

## 6 Conclusions and future work

In the context of airline network and schedule planning, demand forecasting models arouse an increasing interest in order to better understand the underlying travel behavior of passengers. In this paper, an itinerary choice model is developed based on a real dataset

Table 5: Demand indicators for the alternatives for the 3 OD pairs

| | alt. | stops | class | VOT($\frac{€}{h}$) | price elas. | time elas. |
|---|---|---|---|---|---|---|
| **OD1** | 1 | one-stop | E | 19.79 | -2.15 | -0.33 |
| | 2 | one-stop | E | 10.97 | -2.12 | -0.32 |
| | 3 | one-stop | E | 9.22 | -1.97 | -0.41 |
| | 4 | non-stop | E | 8.01 | -1.85 | -0.10 |
| | 5 | non-stop | E | 8.01 | -1.80 | -0.10 |
| | 6 | non-stop | E | 8.01 | -1.94 | -0.10 |
| | 7 | non-stop | B | 21.68 | -1.90 | -0.12 |
| | 8 | non-stop | E | 8.01 | -2.01 | -0.11 |
| | 9 | non-stop | B | 21.68 | -1.86 | -0.11 |
| | 10 | non-stop | E | 8.01 | -2.03 | -0.11 |
| | 11 | non-stop | B | 21.68 | -1.95 | -0.12 |
| | 12 | non-stop | E | 8.01 | -2.01 | -0.11 |
| **OD2** | 1 | one-stop | E | 8.78 | -1.69 | -0.17 |
| | 2 | non-stop | E | 6.86 | -1.37 | -0.06 |
| | 3 | non-stop | E | 6.86 | -2.17 | -0.10 |
| | 4 | non-stop | E | 6.86 | -1.67 | -0.08 |
| | 5 | one-stop | B | 39.81 | -1.93 | -0.32 |
| | 6 | one-stop | E | 21.11 | -2.11 | -0.29 |
| | 7 | one-stop | B | 29.31 | -1.91 | -0.31 |
| | 8 | one-stop | E | 12.29 | -2.08 | -0.29 |
| **OD3** | 1 | one-stop | B | 27.38 | -1.95 | -0.36 |
| | 2 | one-stop | E | 13.60 | -2.14 | -0.33 |
| | 3 | non-stop | E | 10.87 | -1.99 | -0.14 |
| | 4 | non-stop | E | 10.87 | -1.71 | -0.12 |
| | 5 | one-stop | E | 21.41 | -2.16 | -0.29 |
| | 6 | one-stop | E | 11.41 | -2.15 | -0.29 |
| | 7 | non-stop | E | 10.87 | -1.93 | -0.14 |
| | 8 | non-stop | E | 10.87 | -1.88 | -0.14 |
| | 9 | non-stop | E | 10.87 | -1.93 | -0.14 |
| | 10 | non-stop | E | 10.87 | -1.79 | -0.13 |

Table 6: Attributes of the itineraries and the resulting market shares

| OD | price | time of day | market share |
|---|---|---|---|
| A-$B_1$ | 225 | 0 | 0.26 |
| A-$B_2$ | 203 | 1 | 0.44 |
| A-$B'$ | 220 | 0 | 0.30 |

which is aimed to be integrated in a schedule planning model in order to explicitly model supply-demand interactions.

A combined RP/SP dataset is utilized for the estimation of the parameters in order to take the advantage of the elasticity of the SP data. The combination is carried out by constraining a subset of the parameters of the two models to be the same and by introducing a scale parameter for the SP model. As a result, a price elastic demand model is obtained with the help of the combination of the two datasets.

As a future work, the prediction power of the model needs to be analyzed by applying the model on a validation data. The RP data presented in this paper is not very rich in terms of the available explanatory variables. In the existence of a richer dataset the model can be improved.

# References


Belobaba, P., Odoni, A. and Barnhart, C. (2009). *The Global Airline Industry*, first edn, John Wiley & Sons, Ltd, Chichester, UK.

Ben-Akiva, M. and Morikawa, T. (1990). Estimation of travel demand models from multiple data sources, *in* M. Koshi (ed.), *Transportation and Traffic Theory*, Elsevier, New York, pp. 461–476.

Bierlaire, M. and Fetiarison, M. (2009). Estimation of discrete choice models: extending BIOGEME, *Swiss Transport Research Conference (STRC)*.

Carrier, E. (2008). *Modeling the choice of an airline itinerary and fare product using booking and seat availability data*, PhD thesis, Massachusetts Institute of Technology.

Coldren, G. M. and Koppelman, F. S. (2005). Modeling the competition among air-travel itinerary shares: GEV model development, *Transportation Research Part A: Policy and Practice* **39**(4): 345–365.

Coldren, G. M., Koppelman, F. S., Kasturirangan, K. and 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.

Garrow, L. A. (2010). *Discrete Choice Modelling and Air Travel Demand: Theory and Applications*, Ashgate Publishing: Aldershot, United Kingdom.

Gramming, J., Hujer, R. and Scheidler, M. (2005). Discrete choice modelling in airline network management, *J. Appl. Econ* **20**: 467–486.

Koppelman, F. S., Coldren, G. M., Kasturirangan, K. and Parker, R. A. (2008). Schedule delay impacts on air-travel itinerary demand, *Transportation Research Part B: Methodological* **42**: 263–273.

Louviere, J. J., Meyer, R. J., Bunch, D. S., Carson, R., Dellaert, B., Hanemann, W. M., Hensher, D. and Irwin, J. (1999). Combining sources of preference data for modeling complex decision processes, *Marketing Letters* **10**(3): 205–217.

Wen, C.-H. and Lai, S.-C. (2010). Latent class models of international air carrier choice, *Transportation Research Part E: Logistics and Transportation Review* **46**: 211–221.

# A Appendix - Additional estimation results

Table 7: Estimated parameters based on the SP data

| Parameters | Estimated value | t-test |
|---|---|---|
| $\beta^{E,NS}_{\text{price}}$ | -9.63 | -24.05 |
| $\beta^{B,NS}_{\text{price}}$ | -8.50 | -10.21 |
| $\beta^{E,S}_{\text{price}}$ | -9.37 | -24.89 |
| $\beta^{B,S}_{\text{price}}$ | -8.51 | -10.63 |
| $\beta^{B-OP}_{\text{price}}$ | 3.52 | 3.52 |
| $\beta^{E,NS}_{\text{time}}$ | -0.439 | -4.91 |
| $\beta^{B,NS}_{\text{time}}$ | -0.456 | -2.99 |
| $\beta^{E,S}_{\text{time}}$ | -0.328 | -4.23 |
| $\beta^{B,S}_{\text{time}}$ | -0.361 | -2.76 |
| $\beta^{E}_{\text{morning}}$ | 0.122 | 1.28* |
| $\beta^{B}_{\text{morning}}$ | 0.341 | 2.10 |

(* Statistical significance < 90%)

Table 8: Estimated parameters for the model with joint RP and SP data

| Parameters | Estimated value | scaled value for SP | t-test |
|---|---|---|---|
| $\beta^{E,NS}_{\text{price}}$ | -2.23 | -9.63 | -3.48 |
| $\beta^{B,NS}_{\text{price}}$ | -1.97 | -8.49 | -3.64 |
| $\beta^{E,S}_{\text{price}}$ | -2.17 | -9.37 | -3.48 |
| $\beta^{B,S}_{\text{price}}$ | -1.97 | -8.49 | -3.68 |
| $\beta^{B-OP}_{\text{price}}$ | 0.813 | 3.52 | 2.91 |
| $\beta^{E,NS}_{\text{time}}$ | -0.102 | -0.440 | -2.85 |
| $\beta^{B,NS}_{\text{time}}$ | -0.104 | -0.449 | -2.43 |
| $\beta^{E,S}_{\text{time}}$ | -0.0762 | -0.329 | -2.70 |
| $\beta^{B,S}_{\text{time}}$ | -0.0821 | -0.354 | -2.31 |
| $\beta^{E}_{\text{morning}}$ | 0.0283 | 0.122 | 1.21* |
| $\beta^{B}_{\text{morning}}$ | 0.0790 | 0.341 | 1.86 |

(* Statistical significance < 90%)

# B Appendix - Estimation with 24 OD pairs from the RP dataset

We provide the estimation results with 24 OD pairs of RP data is included in the mixed RP/SP dataset. There are 165 alternative itineraries in total serving 5503 passengers between these 24 OD pairs.

In Table 9 we see the results for RP and SP observation obtained with the joint estimation. The presented results are in this case are significantly different compared to the estimation results from the SP data (Table 7). In Table 10 we present the resulting demand elasticities and value of time for OD1 when the estimation is carried out with 24 OD pairs. It is observed that the elasticity of demand is reduced significantly compared to the results provided in Table 5. The demand model parameters in this case do not reflect the behavior of passengers and when integrated into the planning model the price of the itineraries are allowed to increase unrealistically because of the inelasticity.

Table 9: Estimated parameters for the model with joint RP (24 OD pairs) and SP data

| Parameters | Estimated value | scaled value for SP | t-test |
|---|---|---|---|
| $\beta_{\text{price}}^{E,NS}$ | -1.28 | -9.57 | -24.73 |
| $\beta_{\text{price}}^{B,NS}$ | -1.16 | -8.64 | -12.46 |
| $\beta_{\text{price}}^{E,S}$ | -1.25 | -9.32 | -25.66 |
| $\beta_{\text{price}}^{B,S}$ | -1.17 | -8.71 | -13.17 |
| $\beta_{\text{price}}^{B-OP}$ | 0.493 | 3.68 | 4.00 |
| $\beta_{\text{time}}^{E,NS}$ | -0.060 | -0.445 | -5.07 |
| $\beta_{\text{time}}^{B,NS}$ | -0.072 | -0.534 | -3.72 |
| $\beta_{\text{time}}^{E,S}$ | -0.045 | -0.333 | -4.39 |
| $\beta_{\text{time}}^{B,S}$ | -0.0058 | -0.429 | -3.350 |
| $\beta_{\text{morning}}^{E}$ | 0.0154 | 0.115 | 1.21* |
| $\beta_{\text{morning}}^{B}$ | 0.0414 | 0.309 | 2.02 |

(* Statistical significance < 90%)

Table 10: Demand indicators for OD1 when estimated with 24 ODs

| | alt. | stops | class | VOT($\frac{€}{h}$) | price elas. | time elas. |
|---|---|---|---|---|---|---|
| **OD1** | 1 | one-stop | E | 20.16 | -1.23 | -0.19 |
| | 2 | one-stop | E | 11.17 | -1.22 | -0.19 |
| | 3 | one-stop | E | 9.38 | -1.13 | -0.25 |
| | 4 | non-stop | E | 8.15 | -1.06 | -0.06 |
| | 5 | non-stop | E | 8.15 | -1.03 | -0.06 |
| | 6 | non-stop | E | 8.15 | -1.11 | -0.06 |
| | 7 | non-stop | B | 25.33 | -1.12 | -0.08 |
| | 8 | non-stop | E | 8.15 | -1.15 | -0.06 |
| | 9 | non-stop | B | 25.33 | -1.10 | -0.08 |
| | 10 | non-stop | E | 8.15 | -1.16 | -0.06 |
| | 11 | non-stop | B | 25.33 | -1.15 | -0.08 |
| | 12 | non-stop | E | 8.15 | -1.16 | -0.06 |