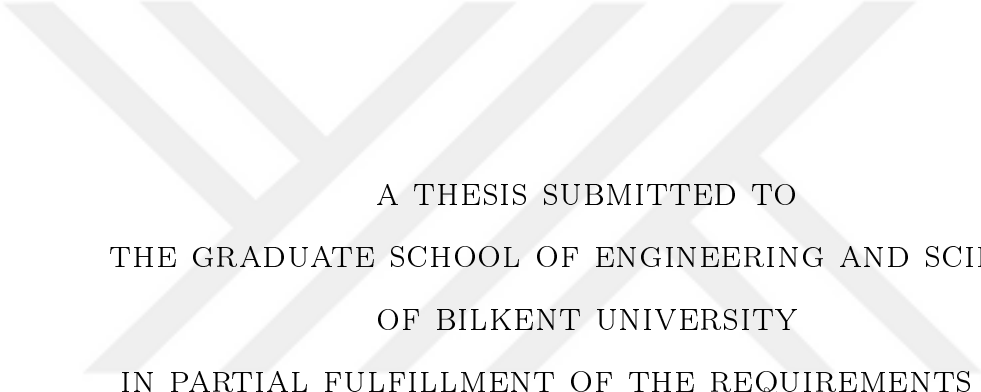


MISSED FLIGHT COVER DESIGN



A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE
OF BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF
MASTER OF SCIENCE
IN
INDUSTRIAL ENGINEERING

By
Beyza Çelik
July 2019

MISSED FLIGHT COVER DESIGN

By Beyza Çelik

July 2019

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.



Savaş Dayanık(Advisor)

Semih Onur Sezer

Alper Şen

Approved for the Graduate School of Engineering and Science:

Ezhan Karaşan
Director of the Graduate School

ABSTRACT

MISSED FLIGHT COVER DESIGN

Beyza Çelik
M.S. in Industrial Engineering
Advisor: Savaş Dayanık
July 2019

Missed flight cover is an option with a price and validity period and is a source of ancillary revenues for the airline companies and helps passengers, who missed their flights, resume their journeys at reduced costs. We study optimal price and validity period of this option to allow a passenger to use missed flight fare towards the purchase of a future airline ticket. Our objective is to maximize the expected ancillary revenues of the airline. The possible actions of passengers are described with a probabilistic graphical model. Within that model, passenger's decision to buy the option and to resume the journey after a missed flight are described with separate hierarchical Bayesian mixed logit regression models. To estimate the parameters of those mixed logit models, an individualized Bayesian choice-based conjoint experiment is designed. In this experiment, each choice set is optimally picked so as to maximize the expected Kullback-Leibler divergence between subsequent posterior distributions of individualized part-worths. The posterior distributions of unknown model parameters, particularly, individualized part-worths, are calculated with a hybrid Markov Chain Monte Carlo (MCMC) algorithm. We developed an R-Shiny online survey web application for six different individualized choice experiments (buy or not buy an option for leisure and business travel, resume or not resume a missed leisure or business flight with or without an option) and collected responses of over 300 individuals. Using the MCMC samples of individual part-worths from their posterior distributions, we simulated the market. We searched and found an option design that maximized the average net revenue of the airline over the simulated runs of the market.

Keywords: Missed flight cover, revenue management, probabilistic graphical models, discrete choice models, Markov Chain Monte Carlo.

ÖZET

YANMAZ BİLET OPSİYONU TASARIMI

Beyza Çelik
Endüstri Mühendisliği, Yüksek Lisans
Tez Danışmanı: Savaş Dayanık
Temmuz 2019

Yanmaz bilet opsiyonu, havayolu şirketlerinin yan gelirlerini arttırırken, küçük kusurlardan dolayı uçaklarını kaçıran yolcuların mâli külfetlerini azaltarak yolculuklarına devam etmelerini sağlayan, ücreti ve geçerlilik süresi olan yeni bir sözleşmeli üründür. Amacımız, beklenen net geliri, opsiyonun fiyatı ve geçerlilik süresi üzerinden eniyileyerek, opsiyonun en büyük beklenen net gelirini, en iyi opsiyon fiyatı ve süresini bulmaktır. İlk olarak, net gelirin beklenen değerini hesaplayabilmek için yolcuların muhtemel davranışları olasılıksal grafik modeller yardımıyla tanımladık. Yolcuların opsiyonu alma ve uçaklarını kaçırdıktan sonra yolculuklarına devam etme kararlarını ise iki ayrı hiyerarşik Bayesyen karışık logit regresyon modeli yardımıyla modelledik. Daha sonra, karışık logit modellerinin parametrelerini tahmin edebilmek için, bireyselleştirilmiş Bayesyen istatistiksel sonlu seçenekler deneyleri tasarladık. Bu deneyleri, her soru seti, bireysel model katsayılarının önsel ve sonsal dağılımları arasındaki, beklenen Kullback-Leibler uzaklığı eniyileyecek şekilde tasarladık. Bilinmeyen model parametrelerinin, başka bir deyişle bireysel model katsayılarının, sonsal dağılımlarını hibrit bir Markov Zinciri Monte Carlo (MZMC) algoritması kullanarak hesapladık. Son olarak, altı farklı bireyselleştirilmiş sonlu seçenekler deneyi (iş ve iş dışı yolculuklar için opsiyon satın alma ya da almama, iş için olan ya da iş dışı olan, opsiyonlu ya da opsiyonsuz, kaçırılmış bir uçuşa devam etme ya da etmeme) için bir R-Shiny çevrimiçi anket web uygulaması geliştirip, bu uygulamayı kullanarak 300'ün üzerinde kişiden veri topladık. Uygulama yardımıyla topladığımız verileri kullanarak elde ettiğimiz, kişisel katsayıların sonsal dağılımlarından gelen MZMC örneklerini, piyasayı simüle etmek için kullandık. En sonunda, simüle edilmiş piyasa değerleri üzerinden havayolu şirketinin beklenen net gelirini eniyileyen opsiyon tasarımını bulduk.

Anahtar sözcükler: Yanmaz bilet opsiyonu, gelir yönetimi, olasılıksal grafik modeller, sonlu seçenekler modelleri, Markov Zinciri Monte Carlo.

Acknowledgement

First and foremost, I would like to thank to my advisor Prof. Savaş Dayanık. I am deeply grateful to him for his patience, support and guidance throughout my study. Without his encouragement and guidance, this thesis would not be possible. I consider myself extremely lucky to have a chance to work with him.

I am grateful to Assoc. Prof. Alper Şen and Assoc. Prof. Semih Onur Sezer for their valuable time to read and review this thesis.

I would like to acknowledge that this research is supported by grant 118M415 of Program 1001 of TUBITAK, The Scientific and Technological Research Council of Turkey.

Finally, I would like to thank my family for their endless love and support. I feel extremely lucky to have them.

Contents

1	Introduction	1
2	Problem Definition	4
3	Model	7
4	Preliminary Analysis	16
5	Revisiting Option Purchase and Resuming The Missed Flight Models	33
5.1	Hierachical Bayesian Mixed Logit Regression Model	36
5.2	Posterior Distributions	39
6	Choice-Based Conjoint Experiment	45
6.1	Attributes and Levels	46
6.2	Choice Set Design	47
6.3	Implementation	49

7 Numerical Analysis	53
7.1 Exploring Data	55
7.2 Generative Process	58
7.3 Results For Expected Net Revenue	62
8 Conclusion	66
A Problem Definitions for Different MCMC samplers	71
B Online Survey Demographic Queries	73
C Online Survey Query Samples	76
D The Reference Respondent	80

List of Figures

3.1	The probabilistic graphical models that define the joint probability distribution of the random variables used in the design of missed flight cover (a) in the presence (b) absence of option in the market. (b).	9
4.1	Preliminary Survey	17
4.2	Finding ticket price probability density functions from the survey data. The comparison of ticket prices for leisure and business travels by using kernel density estimation (left), normal mixture density estimation of ticket price density functions (blue) for leisure (center) and business travels (right).	19
4.3	The probability that a passenger purchases the missed flight cover	20
4.4	Decision tree representation of the simplified model	21
4.5	The expected net revenue per passenger generated by the option (left) and the optimal option price (right) are shown with level and contour plots.	21

4.6	Means of maximum expected net option revenue per passenger (left) and optimal option price (right) over 1,000 bootstrapped samples of market survey data are shown by level and contour graphs.	23
4.7	Variation in maximum expected net revenue per passenger	25
4.8	Variation in optimal option prices	26
4.9	Expected net option revenue function on which the maximum values and corresponding optimal option prices are marked.	27
4.10	Beta probability distribution fits for different bootstrapped samples	28
4.11	After beta distribution is fitted to the option purchasing probability, the expected net option revenue functions are now smooth function of option prices. The maximum net revenues and the corresponding optimal option prices are marked with dashed lines.	29
4.12	After beta distribution fitted, means of maximum expected net option revenues per passenger (left) and optimal option prices (right) over 1,000 bootstrapped samples of market survey data are shown by level and contour graphs.	30
5.1	DAG model for the mixed logit model coefficients	38
6.1	Implementation of survey	50
6.2	An instance of MCMC resampling process	52
7.1	Breakdown of the number of respondents into the dates that survey stays online and boxplots of survey durations. Two vertical lines were mark 5 and 7 minutes.	56

7.2	Expected net revenue generated by a typical passenger when the missed flight cover is present for different passenger no-show rates	63
B.1	Screenshot of the queries for gender, age and income information .	74
B.2	Screenshot of the query for the fraction business travels	74
B.3	Screenshot of the query for the average minimum and maximum ticket prices for business travels and corresponding purchase times of those tickets	75
B.4	Screenshot of the query for the average minimum and maximum ticket prices for leisure travels and corresponding purchase times of those tickets	75
C.1	Screenshot of a query for the optionLeis problem	77
C.2	Screenshot of a query for the optionBus problem	77
C.3	Screenshot of a query for the resumeLeis problem	78
C.4	Screenshot of a query for the resumeBus problem	78
C.5	Screenshot of a query for the resumeLeisOpt problem	79
C.6	Screenshot of a query for the resumeBusOpt problem	79

List of Tables

4.1	Summary of survey data	18
4.2	The maximum expected net revenue per passenger generated by the option for different values of r and Δp_L	22
4.3	The optimal option price for different values of r and Δp_L	22
4.4	Means (and standard deviations) of maximum expected net option revenue per passenger over 1,000 bootstrapped samples of market survey data	24
4.5	Means (and standard deviations) of optimal option price over 1,000 bootstrapped samples of market survey data	24
4.6	After option purchase probabilities are smoothed with excess beta distribution function, the means (and standard deviations) of maximum expected net option revenues per passenger over 1,000 bootstrapped samples of market survey data	31
4.7	After option purchase probabilities are smoothed with excess beta distribution function, the means (and standard deviations) of optimal option prices over 1,000 bootstrapped samples of market survey data	32

7.1	Attribute level scales of option purchase and flight resume probability functions	54
7.2	Descriptions of covariates	55
7.3	Gender versus income	57
7.4	Age versus income	57
7.5	Statistics for the covariates of passengers who use some of their flights for business purposes (TP = Ticket Price, PT = Purchase Time)	57
7.6	Statistics for the covariates of passengers who use some of their flights for leisure purposes (TP = Ticket Price, PT = Purchase Time)	58
7.7	Optimal designs and corresponding expected net revenue values for every passenger no-show rates	64
A.1	Descriptions of six problems modeled in the survey	71
D.1	The reference respondent demographics for each trip type and their scales	80

Chapter 1

Introduction

Missed flight cover (option) is a contract with a price and validity period and is a source of ancillary revenues for the airline companies and helps passengers, who missed their flights, resume their journeys at reduced costs. In Europe, it has been offered by some budget airlines, such as easyJet and Vueling since 2012. Unlike a travel insurance, this contract does not demand the passengers to submit a document for a valid excuse, but only requires that they show up within four hours of the departure of the missed flight at the airline ticket desk in the same airport. Contract holder has the right to purchase an empty seat on the next available airplane departing before the validity period of the contract ends by paying only the difference between the old and new ticket prices.

Is the missed flight cover profitable for every airline company? What is the maximum expected net revenue per passenger for the airline company? What should the best cover option price and validity period be to attain the maximum expected net revenue per passenger? Those questions have not been addressed in the literature, yet. The goal of our study is to fill this gap and address all of those questions. In this way, we would like to design the most profitable missed flight cover. The gist of our problem is to determine and model precisely all cash flows to the airlines to be generated by the contract. This task is split into three parts:

- (1) What factors can cause a passenger to purchase the contract and how do these factors determine the likelihood of a sale? The immediate observable factors that come to mind are the flight class and ticket price of the passenger, the price and validity period of the contract. We try to establish the relationship between sales probability and those factors with a discrete choice model.
- (2) How does the contract change a typical passenger's probability of resuming her journey after she misses her flight? On the one hand, business travellers feel obligated to resume their journey after missing their flights. Therefore, the contract is unlikely to boost the likelihood of that. On the other hand, leisure travellers will be highly discouraged from resuming their journey by the high prices of seats in the airplanes departing on the same or next day. The contract will then increase significantly the likelihood that an leisure traveller will resume her journey by lowering the cost of the journey to the price difference between old and new tickets. Thus the contract has two benefits: i) it is itself a source of ancillary revenues, ii) empty seats on airplanes departing soon will generate revenues when they are bought by the leisure travellers who buy them at discount, thanks to the contract after they miss their original flights. In this part of the study, we capture the likelihood of resuming the journey after a missed flight with a separate discrete choice model.
- (3) How do we select the best of all discrete choice models that we can build for the purchase of a contract or a new seat after missing a flight? Firstly, we design a discrete choice experiment after carefully studying their design theory. We formulate survey questions that can help best to reveal the variations in passenger choices. Thanks to ever closing gap between bus and airplane ticket prices, almost everyone around us can afford travelling by air. Therefore, it is not difficult to collect the choice data we need to pick the best choice models from the potential passengers by means of an online survey.

In the remainder of study, we bring the pieces together to get a mathematical

expression for the net revenue per passenger. Finally, we maximize the expected net revenue per passenger to get the optimal contract price and validity period. To calculate the expectation, we need the joint probability distribution of all of random factors, which are typically not statistically independent. We use a probabilistic graphical model to describe the joint probability distribution of the factors. This allows us to efficiently calculate the expectations without making any oversimplified assumptions and find optimal design parameters of missed flight cover.

The remainder of this thesis is organized as follows. In Chapter 2, problem definition and discussion for the factors that may affect the likelihood of a passenger to purchase the missed flight cover are given. In Chapter 3, the probabilistic graphical model and the expected net revenue generated by a typical passenger when the missed flight cover is present are described. In Chapter 4, numerical results based on aggregate data are presented. In Chapter 5, we revisit the probabilities of a passenger to purchase the missed flight cover and to resume a missed flight. Then we unify discrete choice and probabilistic graphical models to estimate those probabilities accurately. In Chapter 6, the design and the implementation of the individualized choice-based conjoint experiment are explained. Finally in Chapter 7, the conjoint choice data are explored, the optimal option design and the maximum expected net revenue are calculated over the simulated runs of the market based on the generative process of its graphical model.

Chapter 2

Problem Definition

In this thesis, we work on the most profitable design of a product which is called as *missed flight cover*. This product increases the ancillary revenues of airline companies meanwhile enables passengers resume their journey at lower costs in the case of missing their flights. The cover option can be purchased by the passengers during booking their flights at some small fee. When a passenger with the option misses her flight, the amount that she paid for the airline ticket will be insured for a certain period of time. If the passenger wants to continue traveling in the same direction any time during that period, then the insured original amount will be deducted from her new airline ticket price.

The price and validity period of the option are two important decision variables to be determined during the design of the product. They directly affect the willingness of passengers to buy the product thereby the profitability of airline company. The main purpose of our study is to determine the price and validity period of the option that provide the highest profit for the airline company. In addition, the revenue per passenger generated by the most profitable option design and the change rate in the passengers who resume their journey (buy their new tickets at discount thanks to the option) are also calculated. Thus, an airline company not only can find out the optimal price and validity period of the cover option but also measure the contribution of new product to its ancillary revenues

and the increased ticket sales to its main revenues.

In Europe, a slightly different form of missed flight cover has been offered by some budget airlines, such as easyJet and Vueling since 2012. After passenger selecting the cover option during the ticket purchase process, if she misses her flight, and within four hours of the departure of her missed flight, she shows up at the airline ticket desk in the same airport, then the airline has two offers: The passenger can either book a ticket on the next available airplane departing within 24 hours or request a refund. She is not required to give an excuse for missing her flight, and this makes the missed flight cover appealing.

Travel insurance can be an alternative product for distressed passengers due to the possibility of missing a flight. However, unlike the cover option, travel insurance does not provide an immediate replacement of an airline ticket to passenger. In this case, passenger either terminates her journey or purchase a high price ticket from one of the airplanes departing on the same or next day. A travel insurance insures a passenger against only qualifying events (e.g., severe health problems or death of a close relative), which is typically underwritten by a third party insurer and requires the passenger to document her eligibility with a health report or death certificate. This type of travel insurance does not provide the passenger a coverage if the flight is missed, for example, due to a miscalculated travel time to the airport.

For a successful option design, it is necessary to correctly model the behavior patterns of passengers. We distinguish the choices and behavior of each passenger on leisure and business travel. Most airlines offer multifare classes to take advantage of those distinct behaviors. To address the passenger needs on their leisure and business travels, each fare class is designed to include different sets of attributes such as baggage allowance, seat preference, food, penalties for changes in the ticket, etc. While the most influential factors affecting business passengers were reported as reliability, punctuality, seating comfort and schedules, for leisure passengers, price was the most important factor [1]. Therefore, the lowest fare class appeals mostly to leisure passengers whereas the higher fare classes are

preferred by business travelers. One reason for the heterogeneity among two passenger classes is that the ticket is generally paid by the firm when the passenger is on a business trip. When passengers pay for their flights, the demand for the airline tickets is more elastic to travel costs than when they do not. Since the demand of business travelers for airline tickets is less elastic to travel cost, their willingness to pay for airline service quality is higher than leisure travelers [2].

Missed flight cover is also one of the airline quality services that offers flexibility to passengers. Therefore, we need to investigate whether introducing the cover option into the market changes the fare class preferences of business travelers. The main difference between a flexible ticket (high fare class ticket) and missed flight cover is that the cover option is not valid for any changes intended before the departure time. For example, a passenger with the cover option cannot request a seat on an earlier flight due to an early finished meeting, but a passenger with flexible ticket can. Also the cover option has a restriction that passenger should be in the airport within four hours of the departure of the missed flight, which makes it an inappropriate substitute for a flexible ticket. Even if a passenger decides to use the cover option to change her flight before the departure of her original flight, she still needs to show up in the airport within four hours of the departure. Since business travelers are sensitive to travel time, this restriction is disincentive for them to purchase the cover option instead of flexible ticket. Moreover, travelers are willing to pay higher fees for the flexibility of their tickets when they are confident that their company will bear the cost of changing a ticket [3]. Hence, we do not expect that business travelers will change their behavior patterns to purchase a low fare air ticket by renouncing baggage allowance, seat preference, food and ticket flexibility. Therefore, we believe that introducing missed flight cover to the market will not create any shifts in the demand of flexible ticket. This new product will generate new ancillary revenues for the airline without decreasing the existing ones.

Chapter 3

Model

The net revenues of an airline from the sale of the missed flight cover depend on many random variables, random events, and the option price. Those are defined as below.

T : Trip motive, business or leisure travel

P : Airline ticket price

L : Event that passenger misses her flight

R : Event that passenger resumes her missed flight

ΔP : Ticket price difference for the missed and new flights

B : Event that passenger purchases the missed flight cover

p^o : Missed flight cover price

The random revenue generated by a typical passenger is when the missed flight cover is present equals

$$P + p^o 1_B + 1_{L \cap R} (\Delta P + P 1_{B^c}). \quad (3.1)$$

In private communication with the largest domestic airline company yield management department, we learned that their flights are on average 84% full. This leaves a lot of empty seats to be used to further exploit with, e.g., missed flight cover option. We assume that the opportunity to sell empty seats to last-minute passengers or overbooking will not be lost.

In (3.1), 1_A is an indicator random variable of event A. It takes value one if the event happens and zero otherwise. Every passenger who buys a regular ticket from the airline pays the ticket price P . If a passenger purchases the cover option together with her ticket ($1_B = 1$), then she pays the additional fee p^o for the cover.

If a passenger does not miss her flight ($1_L = 0$) or she misses but does not resume her journey ($1_R = 0$), the revenue generated by that passenger is $P + p^o 1_B$. If she misses her flight and resumes her journey, then the generated revenue depends on whether she purchased the cover option or not:

- If the passenger purchased missed flight cover ($1_B = 1$), then the first ticket price will be deduced from her new airline ticket price and she only pays the price difference ΔP between her ticket prices.
- If passenger did not purchase the cover option ($1_B = 0$), then she pays the full price ($\Delta P + P$) of her new ticket.

Thus, the random revenue generated by a typical passenger when the option is present becomes (3.1). Since this is a random variable, it cannot be maximized over p^o so, but its expected value can. To calculate the expected revenue, the joint distribution of random variables P , ΔP , 1_B , 1_L , and 1_R in expression (3.1) should be identified. For the passengers who miss their flights, the decisions of resuming their journeys depend on whether they purchased the cover option or not: since they buy their second tickets at discount thanks to the option, the probability that a passenger with cover option resuming her journey should be higher than that probability for a passenger without the option. The stochastic dependencies

between aforementioned four random variables are expressed in Figure 3.1 with a probabilistic graphical model.

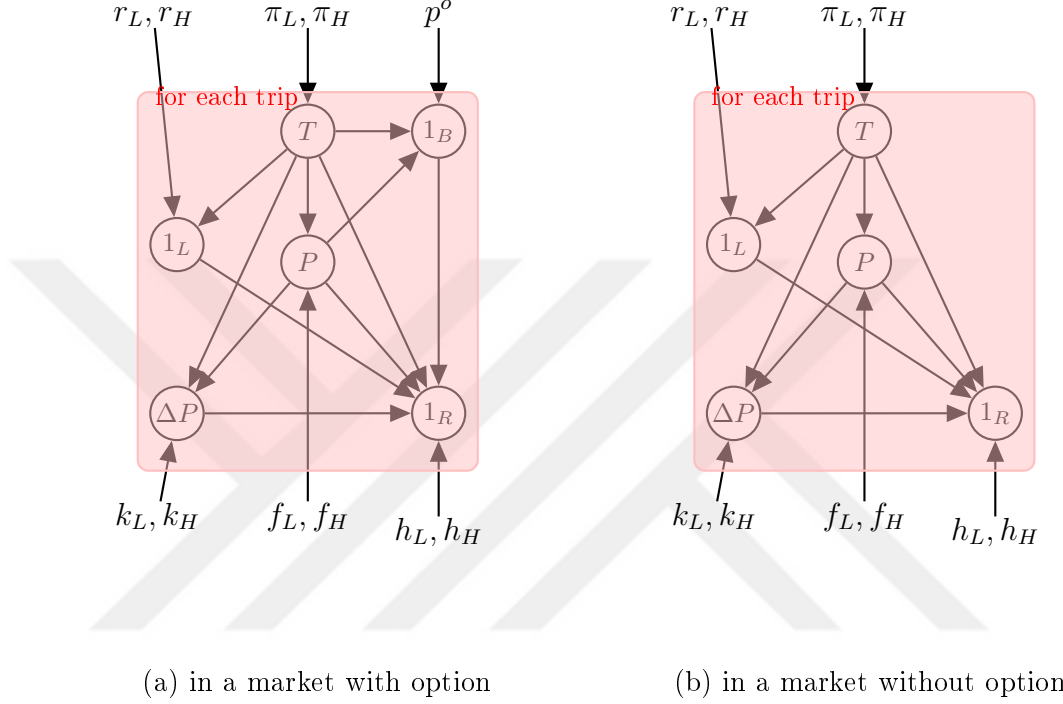


Figure 3.1: The probabilistic graphical models that define the joint probability distribution of the random variables used in the design of missed flight cover (a) in the presence (b) absence of option in the market. (b).

Except option price p^o , the ones outside the box are the parameters of the corresponding random variables that are pointed out with arrows. The arrows in a probabilistic graphical model describe the conditional distributions between random variables. The joint distribution of all random variables equals to the product of conditional distributions between all child-parent nodes existing in the model [4]. According to the model in Figure 3.1a, those conditional probability distributions in a market where missed flight cover option is offered become

$$\begin{aligned}
 \mathbb{P}(T = i) &= \pi_i, \quad i = L, H, \\
 \mathbb{P}(P \in dp \mid T) &= f_T(p) dp, \quad p \geq 0, \\
 \mathbb{P}(\Delta P \in dq \mid T, P) &= k_T(P, q) dq, \quad q \geq 0, \\
 \mathbb{E}(1_B \mid T, P) &= G_T(P, p^o), \\
 \mathbb{E}(1_L \mid T) &= r_T,
 \end{aligned}$$

$$\mathbb{E}(1_R \mid T, 1_B, 1_L, P, \Delta P) = 1_L 1_B H_T(P, \Delta P) + 1_L 1_{B^c} H_T(P, P + \Delta P).$$

We assume that there are two travel modes and those are defined as “ L ” (leisure travel) and “ H ” (business travel). The ticket price for each trip is P , the ticket price difference for a typical passenger who misses her flight is ΔP , and the conditional probability density functions given travel mode T of those random variables

$$\begin{aligned}\mathbb{P}(P \in dp \mid T) &= f_T(p), \\ \mathbb{P}(\Delta P \in dq \mid T, P) &= k_T(P, q),\end{aligned}$$

Also π_L and π_H represent leisure and business travel probabilities for a passenger, whereas r_L and r_H represent the probabilities that the passenger misses her flight when flying for leisure and business, respectively.

When a trip class T , the ticket price P , and option price p^o are given, the probability of a passenger buying the cover option is defined by a general function $P(B = 1 \mid T, P, p^o) = G_T(P, p^o)$. For a passenger who misses her flight, when the price difference ΔP between her new and old tickets is given, another general function

$$H_T(P, 1_B P + \Delta P) = \begin{cases} H_T(P, \Delta P), & \text{if } 1_B = 1, \\ H_T(P, P + \Delta P), & \text{otherwise,} \end{cases} \quad (3.2)$$

defines the probability of a passenger, who is on a T type trip, resuming her journey. Thereby the most important factors of whether a passenger resumes her journey or not after missing her flight become the arguments of this function. We believe that the most suitable forms for those two functions can be found by discrete choice modeling. Simpler forms that we used for the preliminary analysis are presented below.

For each $i = L, R$; $b, \ell, \rho = 0, 1$; $p, q \geq 0$, the graphical model in Figure 3.1a implies the joint probability distribution function

$$\mathbb{P}(T = i, P \in dp, 1_B = b, 1_L = \ell, \Delta P \in dq, 1_R = \rho)$$

$$\begin{aligned}
&= \mathbb{P}(T = i) \mathbb{P}(P \in dp \mid T = i) \mathbb{P}(1_B = b \mid T = i, P = p) \\
&\quad \times \mathbb{P}(1_L = \ell \mid T = i) \mathbb{P}(\Delta P \in dq \mid T = i, P = p) \\
&\quad \times \mathbb{P}(1_R = \rho \mid T = i, 1_B = b, 1_L = \ell, P = p, \Delta P = q) \\
&= \pi_i f_i(p) dp G_i(p, p^o)^b (1 - G_i(p, p^o))^{(1-b)} r_i^\ell (1 - r_i)^{(1-\ell)} k_i(p, q) dq \\
&\quad \times H_i(p, (1-b)p + q)^\rho \left(1 - H_i(p, (1-b)p + q)\right)^{(1-\rho)} 1_{\{\rho \leq \ell\}}.
\end{aligned}$$

According to this distribution, expected revenue per passenger is

$$\mathbb{E}[P + p^o 1_B + 1_{L \cap R}(\Delta P + P 1_{B^c})]. \quad (3.3)$$

We calculate each term of the expression above separately. The expected price of first ticket equals

$$\begin{aligned}
\mathbb{E}P &= \int p \mathbb{P}(P \in dp) = \sum_{i=L,H} \int p \mathbb{P}(T = i, P \in dp) \\
&= \sum_{i=L,H} \int p \mathbb{P}(T = i) \mathbb{P}(P \in dp \mid T = i) = \sum_{i=L,H} \pi_i \int p f_i(p) dp.
\end{aligned}$$

The probability that passenger buys the cover option is

$$\begin{aligned}
\mathbb{P}(B) &= \sum_{i=L,H} \int \mathbb{P}(T = i, P \in dp, 1_B = 1) \\
&= \sum_{i=L,H} \int \mathbb{P}(T = i) \mathbb{P}(P \in dp \mid T = i) \mathbb{E}(1_B \mid T = i, P = p) \\
&= \sum_{i=L,H} \pi_i \int G_i(p, p^0) f_i(p) dp.
\end{aligned}$$

The expected difference between ticket prices of the missed and next flights on the event that original flight is missed and passenger decides to resume her flight is

$$\begin{aligned}
\mathbb{E}1_{L \cap R} \Delta P &= \int q \mathbb{P}(1_L = 1, \Delta P \in dq, 1_R = 1) \\
&= \sum_{i=L,H} \sum_{b=0,1} \int \int q \mathbb{P}(T = i, P \in dp, 1_B = b, 1_L = 1, \Delta P \in dq, 1_R = 1) \\
&= \sum_{i=L,H} \sum_{b=0,1} \int \int q \mathbb{P}(T = i) \mathbb{P}(P \in dp \mid T = i) \mathbb{P}(1_B = b \mid T = i, P = p)
\end{aligned}$$

$$\begin{aligned}
& \times \mathbb{P}(1_L = 1 \mid T = i) \mathbb{P}(\Delta P \in dq \mid T = i, P = p) \\
& \times \mathbb{P}(1_R = 1 \mid T = i, P = p, 1_B = b, 1_L = 1, \Delta P = q) \\
& = \sum_{i=L,H} \pi_i r_i \int \sum_{b=0,1} \left(\int q H_i(p, (1-b)p + q) k_i(p, q) dq \right) G_i(p, p^o)^b \\
& \quad \times (1 - G_i(p, p^o))^{(1-b)} f_i(p) dp \\
& = \sum_{i=L,H} \pi_i r_i \int \left(\int q H_i(p, p + q) k_i(p, q) dq \right) (1 - G_i(p, p^o)) f_i(p) dp \\
& \quad + \sum_{i=L,H} \pi_i r_i \int \left(\int q H_i(p, q) k_i(p, q) dq \right) G_i(p, p^o) f_i(p) dp \\
& = \sum_{i=L,H} \pi_i r_i \int \int q \left(H_i(p, p + q) (1 - G_i(p, p^o)) + H_i(p, q) G_i(p, p^o) \right) \\
& \quad \times k_i(p, q) f_i(p) dq dp,
\end{aligned}$$

The expected price on the events that passenger did not buy the option, missed the flight and decided to resume its flight becomes

$$\begin{aligned}
\mathbb{E} 1_{L \cap R \cap B^c} P &= \int p \mathbb{P}(P \in dp, 1_B = 0, 1_L = 1, 1_R = 1) \\
&= \sum_{i=L,H} \int \int p \mathbb{P}(T = i, P \in dp, 1_B = 0, 1_L = 1, \Delta P \in dq, 1_R = 1) \\
&= \sum_{i=L,H} \int \int p \mathbb{P}(T = i) \mathbb{P}(P \in dp \mid T = i) \mathbb{P}(1_B = 0 \mid T = i, P = p) \\
& \quad \times \mathbb{P}(1_L = 1 \mid T = i) \mathbb{P}(\Delta P \in dq \mid T = i, P = p) \\
& \quad \times \mathbb{P}(1_R = 1 \mid T = i, P = p, 1_L = 1, \Delta P = q) \\
&= \sum_{i=L,H} \pi_i r_i \int p \left(\int H_i(p, p + q) k_i(p, q) dq \right) (1 - G_i(p, p^o)) f_i(p) dp.
\end{aligned}$$

By bringing all the terms together, the expected revenue per passenger which is defined in (3.3) can be rewritten as

$$\mathbb{E}[P + p^o 1_B + 1_{L \cap R}(\Delta P + P 1_{B^c})] = \sum_{i=L,H} \pi_i \int [p + p^o G_i(p, p^o)] f_i(p) dp$$

$$\begin{aligned}
& + \sum_{i=L,H} \pi_i r_i \int \left(\int q H_i(p, p+q) k_i(p, q) dq \right) (1 - G_i(p, p^o)) f_i(p) dp \\
& + \sum_{i=L,H} \pi_i r_i \int \left(\int q H_i(p, q) k_i(p, q) dq \right) G_i(p, p^o) f_i(p) dp \\
& + \sum_{i=L,H} \pi_i r_i \int p \left(\int H_i(p, p+q) k_i(p, q) dq \right) (1 - G_i(p, p^o)) f_i(p) dp.
\end{aligned} \tag{3.4}$$

The optimal option price p^o will be the argument that maximizes the expression (3.4). However, to simplify the expression to some extent and to be sure that the cover option is profitable when the optimal option price is used, we decided to calculate the expected net revenue per passenger and find the optimal option price p^o that maximizes this expected net revenue.

Passenger behaviors in the absence of option is modeled by the graphical model in Figure 3.1b. When a passenger misses her flight, she needs to pay a higher price for the new ticket in the absence of the option and that price may discourage her from resuming her journey. Therefore, the probability of a passenger resuming her journey after missing her flight may decrease. This is captured in $H_i(P, P + \Delta P)$ function introduced by (3.2) on page 10. In the absence of option the expected revenue per passenger is

$$\mathbb{E}^* [P + 1_{L \cap R}(\Delta P + P)]. \tag{3.5}$$

Here, $\mathbb{E}^*$ is the expected value under probability measure $\mathbb{P}^*$ induced by the graphical model in 3.1b describing the market without option. In the presence of option, the joint distribution of the random variables $P, 1_B, 1_L, \Delta P, 1_R$ determines the probability measure $\mathbb{P}$. In the absence of option, the joint distribution of the random variables $P, 1_L, \Delta P, 1_R$ differs from the previous one and determines a probability measure $\mathbb{P}^*$ different than $\mathbb{P}$. According to that probabilistic graphical model, the conditional distributions of random variables in the market without option are

$$\begin{aligned}
\mathbb{P}^*(T = i) &= \pi_i, \quad i = L, H, \\
\mathbb{P}^*(P \in dp \mid T) &= f_T(p) dp, \quad p \geq 0,
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}^*(\Delta P \in dq \mid T, P) &= k_T(P, q) dq, \quad q \geq 0, \\
\mathbb{E}^*(1_R \mid T, P, 1_L, \Delta P) &= 1_L H_T(P, P + \Delta P), \\
\mathbb{E}^*(1_L \mid T) &= r_T.
\end{aligned}$$

In the absence of option, for every $i = L, H; \ell, \rho = 0, 1; p, q \geq 0$, the joint distribution of random variables is

$$\begin{aligned}
&\mathbb{P}^*(T = i, 1_L = \ell, P \in dp, \Delta P \in dq, 1_R = \rho) \\
&= \pi_i f_i(p) dp r_i^\ell (1 - r_i)^{(1-\ell)} k_i(p, q) dq (H_i(p, p + q))^\rho (1 - H_i(p, p + q))^{(1-\rho)} 1_{\{\rho \leq \ell\}}.
\end{aligned}$$

The terms of expected revenue per passenger in the absence of option can be calculated as

$$\begin{aligned}
\mathbb{E}^* P &= \sum_{i=L,H} \int p \mathbb{P}^*(T = i, P \in dp) = \sum_{i=L,H} \int p \mathbb{P}^*(T = i) \mathbb{P}^*(P \in dp \mid T = i) \\
&= \sum_{i=L,H} \pi_i \int p f_i(p) dp, \\
\mathbb{E}^* 1_{L \cap R} \Delta P &= \int q \mathbb{P}^*(1_L = 1, \Delta P \in dq, 1_R = 1) \\
&= \sum_{i=L,H} \int \int q \mathbb{P}^*(T = i, P \in dp, 1_L = 1, \Delta P \in dq, 1_R = 1) \\
&= \sum_{i=L,H} \int \int q \mathbb{P}^*(T = i) \mathbb{P}^*(P \in dp \mid T = i) \mathbb{P}^*(1_L = 1 \mid T = i) \\
&\quad \times \mathbb{P}^*(\Delta P \in dq \mid T = i, P = p) \mathbb{P}^*(1_R = 1 \mid T = i, P = p, 1_L = 1, \Delta P = q) \\
&= \sum_{i=L,H} \pi_i r_i \int \left(\int q H_i(p, p + q) k_i(p, q) dq \right) f_i(p) dp, \\
\mathbb{E}^* 1_{L \cap R} P &= \int p \mathbb{P}^*(P \in dp, 1_L = 1, 1_R = 1) \\
&= \sum_{i=L,H} \int \int p \mathbb{P}^*(T = i, P \in dp, 1_L = 1, \Delta P \in dq, 1_R = 1) \\
&= \sum_{i=L,H} \int \int p \mathbb{P}^*(T = i) \mathbb{P}^*(P \in dp \mid T = i) \mathbb{P}^*(1_L = 1 \mid T = i) \\
&\quad \times \mathbb{P}^*(\Delta P \in dq \mid T = i, P = p) \mathbb{P}^*(1_R = 1 \mid T = i, P = p, 1_L = 1, \Delta P = q) \\
&= \sum_{i=L,H} \pi_i r_i \int p \left(\int H_i(p, p + q) k_i(p, q) dq \right) f_i(p) dp.
\end{aligned}$$

The expected revenue per passenger in the absence of option defined in expression (3.5) becomes

$$\begin{aligned}\mathbb{E}^*[P + 1_{L \cap R}(\Delta P + P)] &= \sum_{i=L,H} \pi_i \int p f_i(p) dp \\ &+ \sum_{i=L,H} \pi_i r_i \int \left(\int q H_i(p, p+q) k_i(p, q) dq \right) f_i(p) dp \\ &+ \sum_{i=L,H} \pi_i r_i \int p \left(\int H_i(p, p+q) k_i(p, q) dq \right) f_i(p) dp.\end{aligned}$$

Finally, the expected net revenue per passenger generated by the option equals the difference between expressions (3.3) and (3.5); namely,

$$\begin{aligned}\mathbb{E}[P + p^o 1_B + 1_{L \cap R}(\Delta P + P 1_{B^c})] - \mathbb{E}^*[P + 1_{L \cap R}(\Delta P + P)] \\ = \sum_{i=L,H} \pi_i \int \left(p^o + r_i \int (q H_i(p, q) - (p+q) H_i(p, p+q)) k_i(p, q) dq \right) \\ \times G_i(p, p^o) f_i(p) dp.\end{aligned}\tag{3.6}$$

The optimal option price p^o is determined so as to maximize the net revenue function in (3.6). If the net revenue generated by a typical passenger when this option is present is positive, then the missed flight cover is profitable for the airline, and the ancillary net revenue generated by this option reaches its maximum expected net value.

In the next chapter, the effect of option price to the airline ancillary revenues is examined in detail with a preliminary analysis. By means of a preliminary market survey with 65 people, some information were gathered about the flight patterns of passengers and their reactions to the missed flight cover. Using gathered information we build our initial model and by means of the model based on aggregate data, the expected revenue per passenger is calculated for different option prices. Among those, the one that generates the maximum expected revenue is selected.

Chapter 4

Preliminary Analysis

To foresee the difficulties in designing the most profitable missed flight cover and to learn how to overcome those difficulties, we conducted a numerical study, using data collected with a simple survey in a related but independent study in 2015. Firstly, randomly selected 117 people are requested to fill out the survey in Figure 4.1. As a part of their senior design course project six undergraduate students from Bilkent University, Industrial Engineering Department designed the survey in Figure 4.1 and used it to collect 53 responses in Esenboğa Airport domestic terminal, 13 responses in Esenboğa Airport international terminal, 18 responses on Bilkent University Campus, and 33 responses from a website called Survey-Monkey. After eliminating incomplete and inconsistent responses, summary of the remaining 65 responses are presented in Table 4.1, where the minimum and maximum, first, second (median) and third quartiles of every variable are shown for each travel type. Among those variables, the data that are gathered from the average airline ticket prices reported on question 5 in Figure 4.1 (6th line in Table 4.1) are used to model the ticket price (P) probability density functions for both leisure (f_L) and business (f_H) motivated travels.

Figure 4.2 shows the distribution of average ticket prices found in the survey. On the left, we see that the kernel density function concentrates on distinct clusters. One plausible explanation is that there are a few destinations of different

1. Gender:	☐ Female	☐ Male
2. Age:	_____	
3. Do you fly for leisure or business?:	☐ Leisure	☐ Business
4. How many times do you fly per year ? :	_____	
5. How much do you pay for an airline ticket on average ? :	_____	TL
6. How far in advance do you book your flight ? :	_____	
7. Do you feel worried about missing your flight ? :		
	☐ Always	☐ Often ☐ Sometimes ☐ Never
8. Have you ever missed your flight ? :	☐ Yes	☐ No
9. Would you like to use your ticket at a later date by paying your fare difference if you miss your flight ? :	☐ Yes	☐ No
10. Are you willing to pay a price when purchasing your ticket to benefit from the above mentioned option ? :	☐ Yes	☐ No
11. If you consider the average price for your airline tickets how much do you think the above option should cost ? :	_____	TL
12. When would you like to continue on your journey after missing your flight ? :		
	☐ Within 24 hours	☐ Other _____

Figure 4.1: Preliminary Survey

distances and ticket prices. Therefore, the ticket prices can be modeled as mixtures of normal density functions. The mixture coefficient, mean and standard deviation for each component of the the mixture can be found with Bayesian Expectation Maximization (BEM) algorithm [5, 6]. The normal density functions for the ticket prices of leisure and business travels, found by BEM method, are shown in the center and right of Figure 4.2, respectively.

When estimating the density functions, we are not required to know the number and positions of their components. For instance, when finding the normal density functions in Figure 4.2, the initial value for component number is determined as 10 which is more than we expect and the initial position for each component is randomly assigned. However, BEM method forces the mixing coefficients of 6 or 7 components to fall to almost zero. In this way, the true component number is estimated as 4 or 3 and the positions of those clusters are determined so as to maximize the posterior probabilities.

The method BEM softens the extreme values in ticket price distributions by combining very small components and this helps the calculated net option revenue to be more stable. The softening degree can be adjusted by changing the prior

Table 4.1: Summary of survey data

		Leisure		Business		
Rate (%)		46		54		
Those who are willing to buy the option (%)		73		86		
Variable	Travel Type	Min	P25	P50	P75	Max
Age	Leisure	18	21	23	26	46
	Business	17	30	35	40	60
Average annual number of flights	Leisure	0	5	6	15	23
	Business	1	10	10	20	100
How far in advance purchased ticket (days)	Leisure	0	7	14	30	90
	Business	1	7	10	16	90
Option price willing to pay (TL)	Leisure	0	1	28	50	150
	Business	0	10	20	50	250
Average ticket price (TL)	Leisure	50	128	175	250	1,500
	Business	80	105	150	250	3,000
Option Price/Average ticket price (%)	Leisure	0	2	10	24	50
	Business	0	7	13	25	67
Preferred option validity period (days)	Leisure	1	1	1	1	1
	Business	0	1	1	1	30

distribution parameters of the mixing coefficients, means and standard deviations. The optimal number for components and mean and standard deviation for each component are determined by repeating the mentioned procedure 20 times and choosing the one with maximum posterior probability among them.

The function $g(\cdot)$ is estimated by using the amounts that the 65 respondents of our survey are willing to pay to purchase the option (question 11, 4th line in Table 4.1), and the average ticket prices of those respondents (question 5, 6th line in Table 4.1). The estimated option purchase probability function, which illustrated in Figure 4.3, is defined as

$$G_L(p, p^o) = G_H(p, p^o) = g(p^o/p),$$

$$\widehat{g(x)} := \frac{1}{65} \sum_{i=1}^{65} 1_{[x, \infty)}(x_i).$$

Here, x_i is the ratio of the i^{th} passenger's appraised value of the option to his/her average ticket price.

According to Figure 4.3, three quarters of passengers are willing to pay at most one eighth of their ticket prices for the option. If the option price is more than

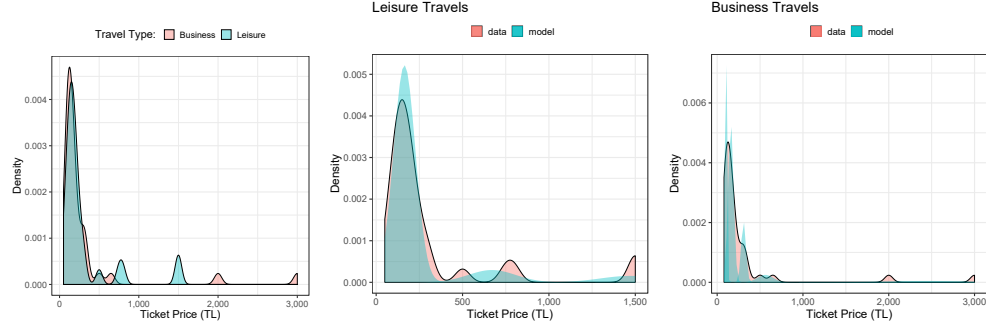


Figure 4.2: Finding ticket price probability density functions from the survey data. The comparison of ticket prices for leisure and business travels by using kernel density estimation (left), normal mixture density estimation of ticket price density functions (blue) for leisure (center) and business travels (right).

one fourth of the ticket price, the fraction of passengers who are willing to buy the option immediately falls to a little over just one quarter.

Before calculating the net revenue of the option, we need to estimate the probability of a passenger resuming her journey after missing her flight; namely, $H_L(p, q)$ and $H_H(p, q)$, for leisure and business travels, respectively. Business travellers, they are generally obligated to resume their journey if they miss their first flights. However, leisure travellers tend not to resume their journey due to high price of a new ticket. In the preliminary study, to be able to observe better the effects of behavior difference between those two travel types, we assume that

$$H_L(p, (1-b)p + q) = \begin{cases} 1, & b = 1 \\ 0, & b = 0 \end{cases}, \quad H_H(p, (1-b)p + q) \equiv 1, \quad \forall b = 0, 1.$$

With the estimation of probabilities, the simplified model which is based on aggregate data is presented in Figure 4.4 with a decision tree representation. The numbers on Figure 4.4 represent different scenarios for a passenger. The profits generated by those scenarios are given as the four terms of the right hand side of (4.1), respectively.

$$\mathbb{E} [P + p^o 1_B + 1_{L \cap R} (\Delta P + P 1_{B^c})] - \mathbb{E}^* [P + 1_{L \cap R} (\Delta P + P)]$$

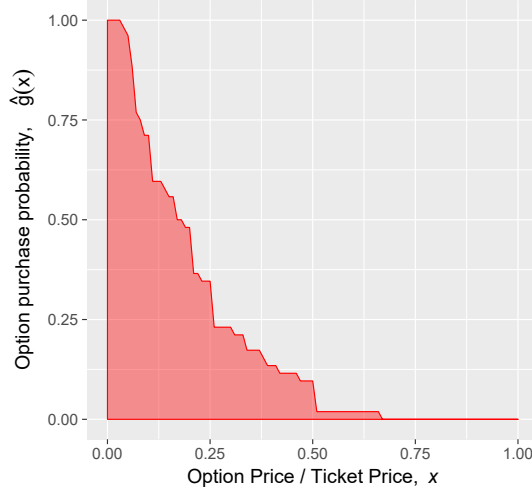


Figure 4.3: The probability that a passenger purchases the missed flight cover

$$\begin{aligned}
&= \pi_L r(p^o + \Delta p_L) \int f_L(p) g\left(\frac{p^o}{p}\right) dp \\
&\quad + \pi_{L(1-r)} p^o \int f_L(p) g\left(\frac{p^o}{p}\right) dp \\
&\quad + (1 - \pi_L)(1 - r) p^o \int f_H(p) g\left(\frac{p^o}{p}\right) dp \\
&\quad + (1 - \pi_L) r \int f_H(p) g\left(\frac{p^o}{p}\right) (p^o - p) dp,
\end{aligned} \tag{4.1}$$

where r is the probability that passenger misses her flight, and Δp_L is the average ticket price difference between the missed and new flights for a leisure travel. We assume that the probability of a passenger missing her flight is the same for both travel types. This assumption enables us to draw expected net revenue values in two dimensions as in Figure 4.5.

Finally, the expected net revenue in (4.1) generated by the missed flight cover is maximized with respect to option price p^o on a discrete grid between 0-60 TL. Both optimal option price p^o and the expected net revenue generated by this option price are calculated. The calculations are repeated for the reasonable values of the parameters, r and Δp_L . Instead of estimating those parameters, the calculations are repeated with some constant values because we believe that the true value of r is really small; therefore, errors in its estimation will be large. Then, the magnitude of the variation in the results for different r values will

be crucial for the estimation method. In addition to this, Δp_L is a parameter controlled by airline company. Therefore, it will be useful to see the effects of variations in Δp_L to the expected net option revenue to determine its effect more accurately.

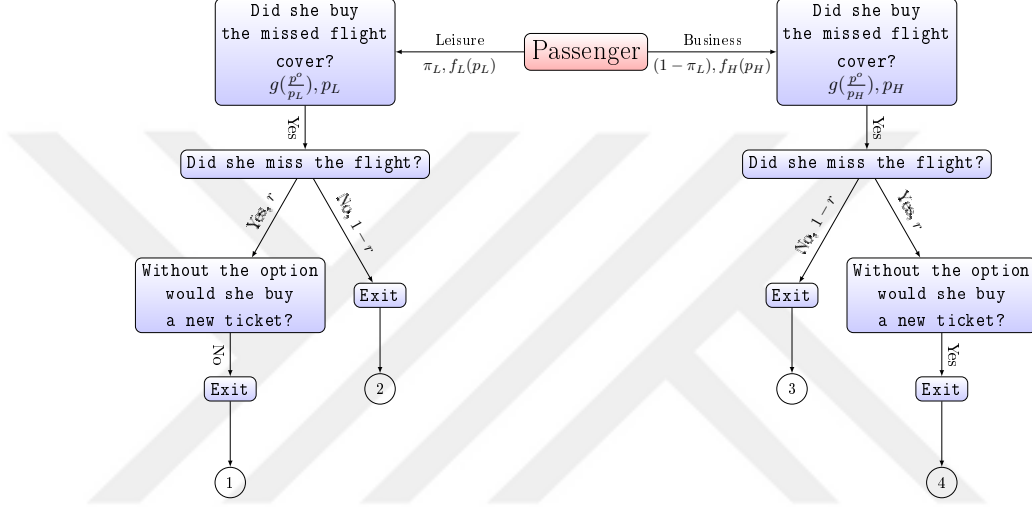


Figure 4.4: Decision tree representation of the simplified model

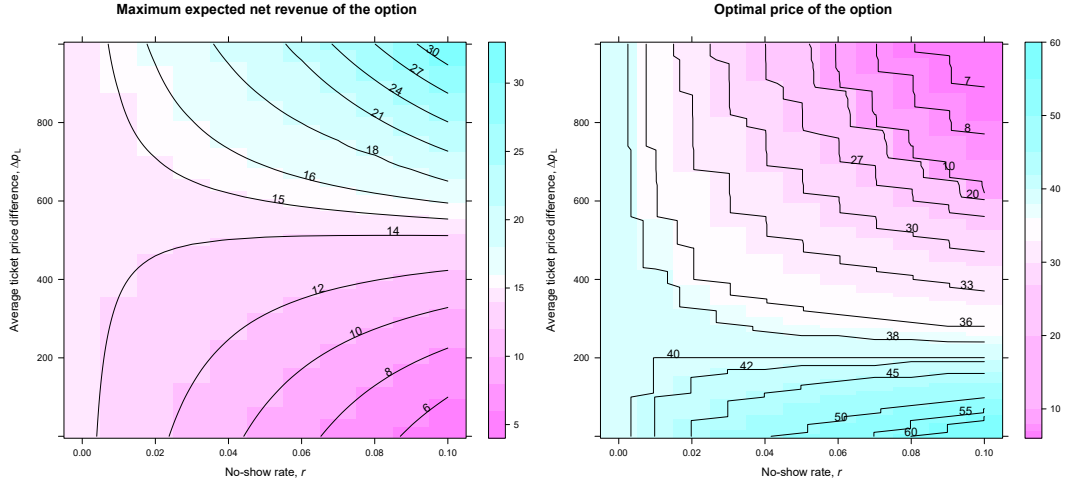


Figure 4.5: The expected net revenue per passenger generated by the option (left) and the optimal option price (right) are shown with level and contour plots.

In Figure 4.5, the expected net revenue generated by missed flight cover (left) and the optimal option price (right) are presented with level and contour graphs. The results for different r and Δp_L parameters can be found in Table 4.2 and 4.3. Expected net option revenue per passenger is between 5-32 TL whereas

optimal option price is between 6-60 TL. Those ranges of expected net revenues and option prices are obtained as we change the probability of missing a flight in the interval between 0%-10% and the ticket price differences for the missed and new flights between 0 – 1000 TL. Expected net option revenue increases with the ticket price difference Δp_L for leisure travels for every fixed flight miss probability r and with r at every fixed Δp_L equal to or greater than 600 TL as well. However, for the values of Δp_L equal to or less than 500 TL, expected net option revenue decreases r at every fixed Δp_L .

Table 4.2: The maximum expected net revenue per passenger generated by the option for different values of r and Δp_L .

Δp_L	r										
	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
1000	14.41	15.24	16.22	17.30	18.48	19.76	21.95	24.42	26.97	29.56	32.22
900	14.41	15.04	15.79	16.63	17.52	18.49	19.65	21.63	23.72	25.86	28.03
800	14.41	14.85	15.37	15.96	16.60	17.28	18.02	18.94	20.54	22.22	23.93
700	14.41	14.65	14.96	15.31	15.70	16.12	16.57	17.05	17.58	18.68	19.92
600	14.41	14.46	14.55	14.68	14.83	15.00	15.19	15.39	15.62	15.86	16.13
500	14.41	14.27	14.16	14.06	13.99	13.92	13.86	13.81	13.78	13.75	13.73
400	14.41	14.08	13.77	13.47	13.17	12.88	12.60	12.32	12.04	11.77	11.50
300	14.41	13.90	13.40	12.89	12.39	11.90	11.40	10.91	10.42	9.92	9.43
200	14.41	13.72	13.04	12.35	11.67	10.98	10.30	9.61	8.93	8.24	7.56
100	14.41	13.55	12.70	11.84	11.00	10.16	9.32	8.49	7.66	6.83	6.00
0	14.41	13.38	12.37	11.38	10.40	9.44	8.49	7.56	6.63	5.73	4.82

Table 4.3: The optimal option price for different values of r and Δp_L .

Δp_L	r										
	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
1000	39	35	31	29	27	23	9	8	7	6	6
900	39	35	32	30	28	25	10	9	8	7	6
800	39	35	33	31	29	27	25	10	9	8	8
700	39	36	33	32	30	29	27	26	24	9	8
600	39	36	35	33	32	30	30	28	27	26	25
500	39	38	35	35	33	33	31	31	30	30	28
400	39	39	36	35	35	35	33	33	33	32	32
300	39	39	39	38	38	36	36	36	35	35	35
200	39	40	40	40	40	40	40	40	40	40	40
100	39	42	42	42	45	45	46	48	48	50	50
0	39	42	45	46	50	52	52	55	60	60	60

After the calculation of maximum expected net revenue and optimal option price we want to determine how much uncertainty lies in our results. Our dataset

consists of only 65 observations, and our results are highly dependent to this limited data; namely, with a different group of passengers, the maximum expected net revenue and the optimal price can change. To express the uncertainty in our results, we construct confidence intervals (CIs) for maximum expected net revenue and optimal option price. By using CI, we get information about plausible ranges for expected net revenues and optimal option prices. To construct confidence intervals without making any assumptions about the distribution of maximum expected net revenue and optimal option price, the bootstrap method is used. By this method, we can determine the distribution of our results without using the mathematical methods of distribution theory. We repeatedly generate artificial data by sampling with replacement from the observed ones. In each bootstrap iteration we calculate the maximum expected net revenue and its corresponding optimal option price and gather the results to study their distributions. The number of bootstrap samples is determined as 1,000.

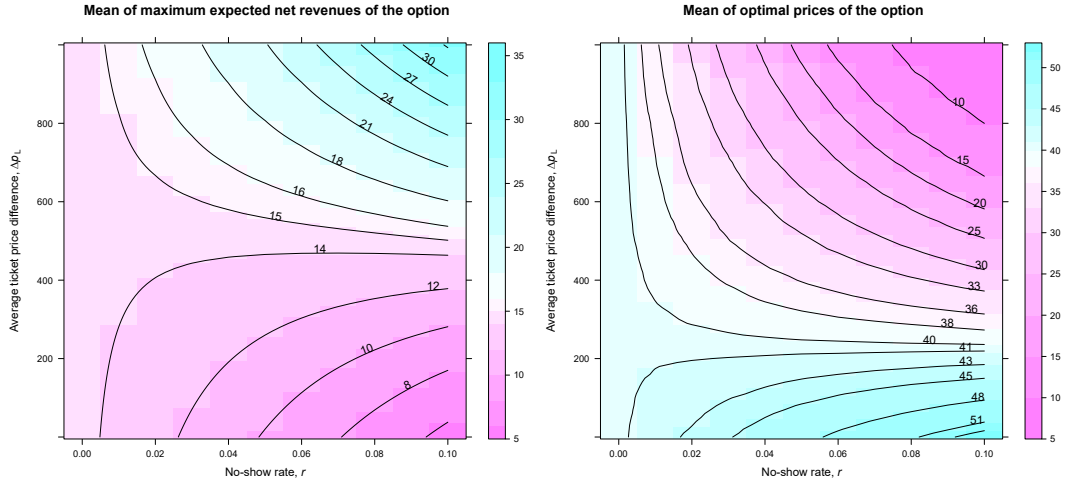


Figure 4.6: Means of maximum expected net option revenue per passenger (left) and optimal option price (right) over 1,000 bootstrapped samples of market survey data are shown by level and contour graphs.

As the next step we construct 95 % confidence intervals for both maximum expected net option revenue and optimal option price. The mean and standard deviation of those statistics are presented in Table 4.4 and 4.5. The variation in maximum expected net option revenue is mostly smaller than the variation in optimal option price. This means that although the optimal values of the option

Table 4.4: Means (and standard deviations) of maximum expected net option revenue per passenger over 1,000 bootstrapped samples of market survey data

Δp_L	r										
	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
1000	14(2)	15(2)	16(2)	18(3)	19(3)	21(4)	23(5)	25(6)	28(7)	31(8)	33(9)
900	14(2)	15(2)	16(2)	17(3)	18(3)	19(4)	21(5)	23(5)	25(6)	27(7)	29(8)
800	14(2)	15(2)	16(2)	16(2)	17(3)	18(4)	19(4)	20(5)	22(6)	23(7)	25(7)
700	14(2)	15(2)	15(2)	16(2)	16(3)	17(3)	17(4)	18(5)	19(5)	20(6)	21(7)
600	14(2)	15(2)	15(2)	15(2)	15(3)	15(3)	16(4)	16(4)	17(5)	17(5)	18(6)
500	14(2)	14(2)	14(2)	14(2)	14(3)	14(3)	14(3)	14(4)	15(5)	15(5)	15(6)
400	14(2)	14(2)	14(2)	14(2)	14(3)	13(3)	13(3)	13(4)	13(4)	13(5)	12(5)
300	14(2)	14(2)	14(2)	13(2)	13(2)	12(3)	12(3)	12(4)	11(4)	11(4)	10(5)
200	14(2)	14(2)	13(2)	13(2)	12(2)	11(3)	11(3)	10(3)	10(4)	9(4)	9(5)
100	14(2)	14(2)	13(2)	12(2)	11(2)	11(3)	10(3)	9(3)	8(4)	8(4)	7(4)
0	14(2)	14(2)	13(2)	12(2)	11(2)	10(2)	9(3)	8(3)	7(3)	6(4)	5(4)

Table 4.5: Means (and standard deviations) of optimal option price over 1,000 bootstrapped samples of market survey data

Δp_L	r										
	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
1000	41(9)	36(8)	33(8)	29(8)	24(10)	18(10)	14(9)	11(8)	9(6)	8(4)	7(4)
900	41(9)	37(8)	34(8)	30(8)	26(9)	22(10)	17(10)	13(9)	11(8)	9(6)	8(5)
800	41(9)	38(9)	35(8)	32(8)	29(9)	25(10)	21(10)	17(10)	14(9)	12(8)	10(7)
700	41(9)	38(9)	35(8)	33(8)	31(8)	28(9)	25(10)	22(11)	19(11)	16(10)	13(9)
600	41(9)	39(9)	36(8)	35(8)	33(8)	31(9)	29(9)	27(10)	24(11)	21(11)	19(11)
500	41(9)	39(9)	38(9)	36(8)	35(8)	33(9)	32(9)	31(9)	29(10)	27(10)	25(11)
400	41(9)	40(9)	39(9)	38(9)	37(9)	36(9)	35(9)	34(9)	33(9)	32(9)	31(10)
300	41(9)	40(9)	40(9)	39(9)	39(9)	39(9)	38(9)	38(9)	37(9)	37(10)	37(10)
200	41(9)	41(9)	41(9)	41(9)	41(9)	41(9)	41(9)	42(9)	42(10)	42(10)	42(10)
100	41(9)	41(9)	42(9)	43(9)	44(9)	44(9)	45(9)	46(9)	46(9)	47(10)	48(10)
0	41(9)	42(9)	43(9)	45(9)	46(9)	47(9)	49(9)	50(9)	51(9)	52(8)	53(8)

price vary over a wide range, the maximum expected net revenue per passenger obtained from this price is stable. As long as the confidence interval for maximum expected net revenue does not include zero we can say that the missed flight cover is profitable. In Table 4.4 and 4.5, when r is greater than or equal to 8 percent, the confidence intervals for the most maximum expected net revenues include zero so in those cases the option is not profitable anymore. However, we do not expect that missed flight probability r of a typical passenger is greater than 8 percent. Hence, this observation has more theoretical than practical importance. For the r values smaller than 8 percent, we can observe that maximum expected net revenue attains positive values.

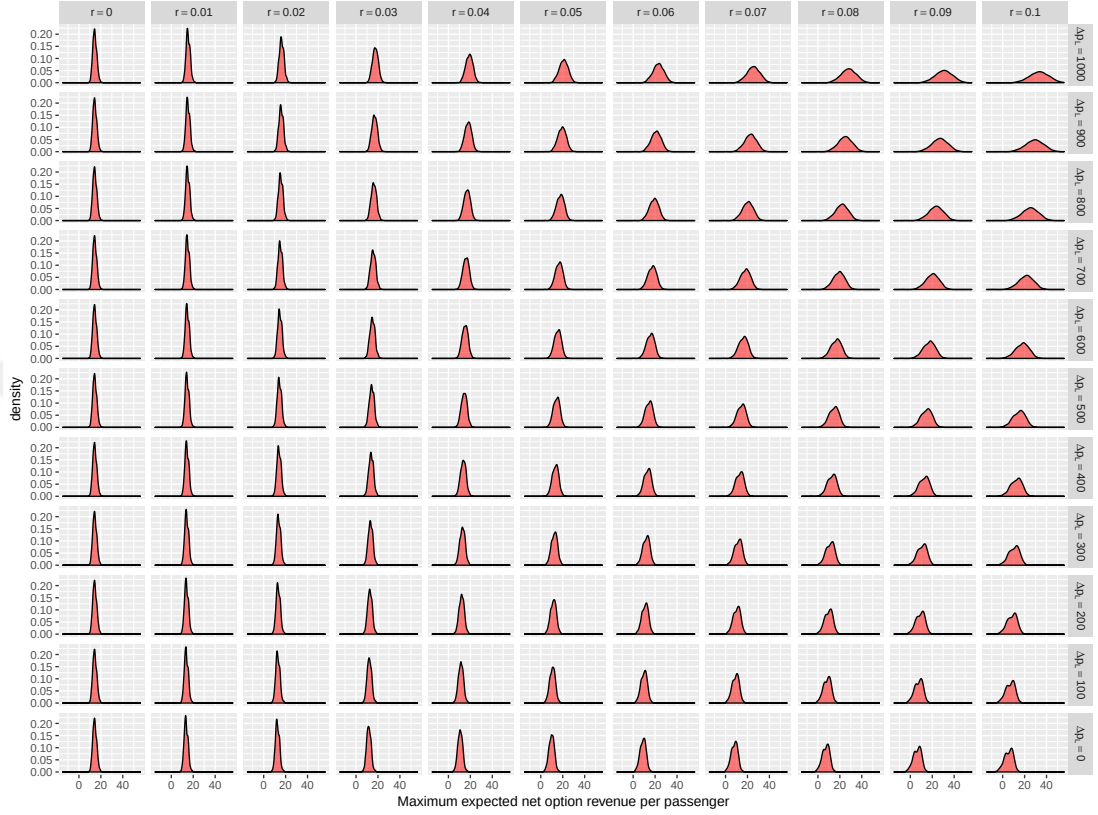


Figure 4.7: Variation in maximum expected net revenue per passenger

In Figure 4.7, for the smaller r values, the distribution of maximum expected revenue per passenger has a taller peak and lighter tails whereas the peaks are getting shorter and tails are getting heavier as both r and Δp_L values increase. In addition to those, we observe that the standard deviation for maximum expected net revenue is also increasing with r and Δp_L values.

In our preliminary study, we assumed that there is a cap (60 TL) on the option price because of either of price regulations or airline's strategic decision to keep the price low. However as illustrated in Figure 4.8, the mass quickly growing at $p^o = 60$ as r increases suggests that the grid search on p^o stops at the edge of search space sub-optimally. To relax this restrictive assumptions we decide to write a function for the expected net option revenue to maximize it using Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm which is an iterative method for solving unconstrained nonlinear optimization problems.

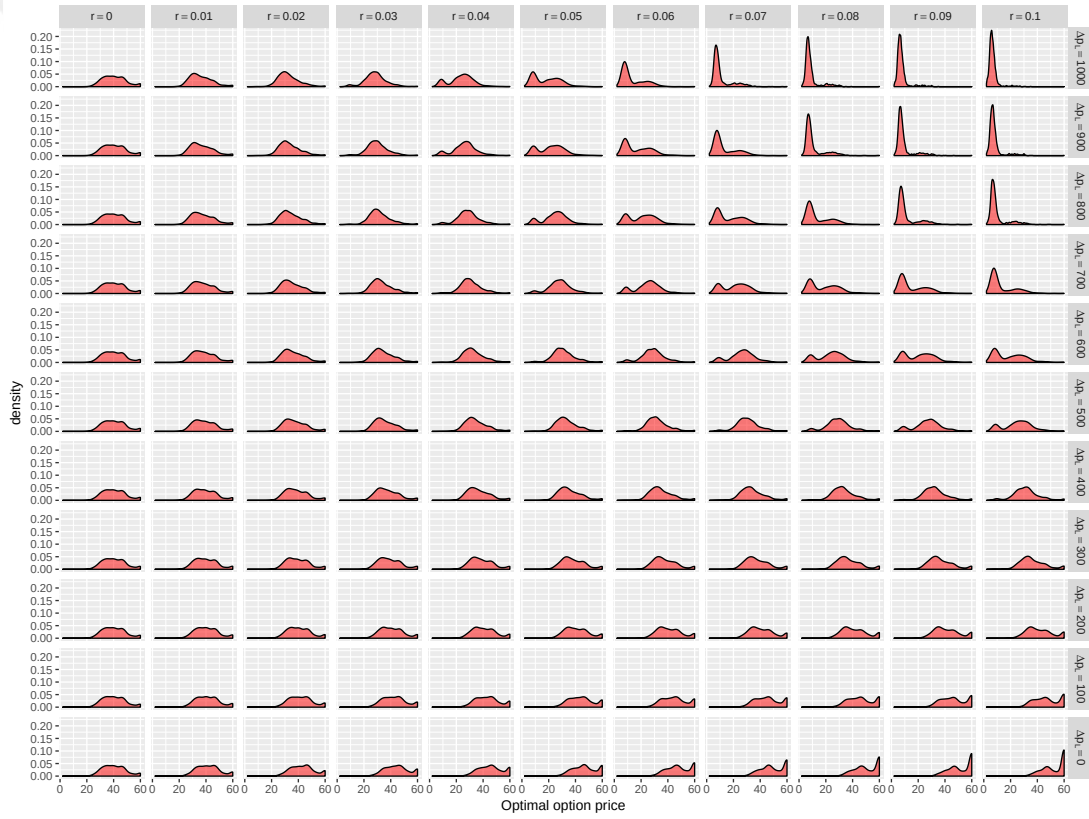


Figure 4.8: Variation in optimal option prices

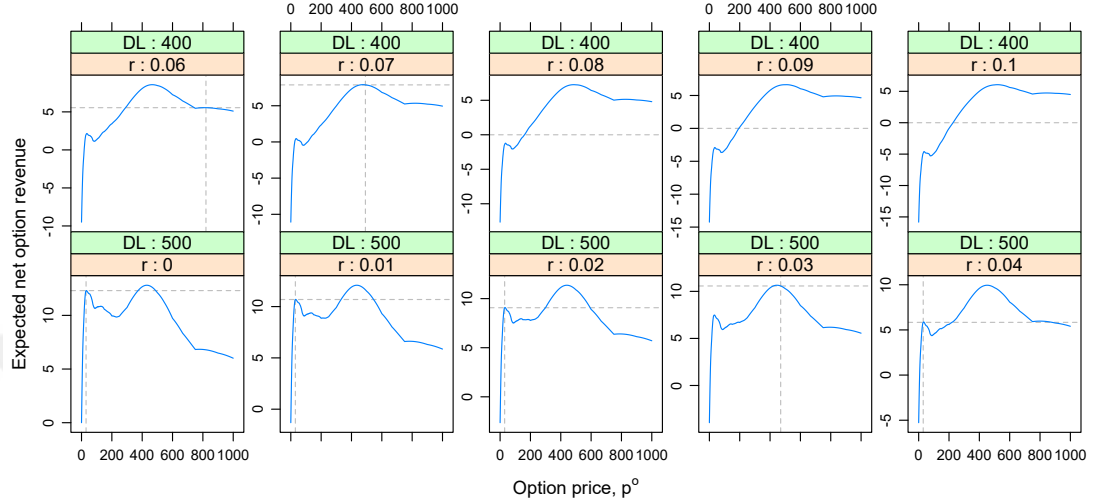


Figure 4.9: Expected net option revenue function on which the maximum values and corresponding optimal option prices are marked.

The BFGS algorithm takes an initial estimate for the optimal value as an input and then searches for better estimates at each iteration. However, it is not guaranteed for this algorithm to converge or to find the global optimum. Therefore, the algorithm may stop at local optima or maximum iteration number may be reached before finding any solution [7]. To avoid those, we combine grid search method with this algorithm. Firstly, we run BFGS with many different initial values in the interval 0-100 TL. Then we select the maximum value among them and take the corresponding option price as the initial estimate for the BFGS algorithm.

In Figure 4.9, we see the resulting expected net option revenue functions are not concave and smooth. Therefore, the search either got stuck in a local optimum or has reached the maximum iteration number before finding any solutions. Furthermore, the non-smooth expected net option revenue function originated from the non-smooth option purchase probability function as shown in Figure 4.3, may cause BFGS to get trapped at a local maximum.

We need to fit a smooth non-increasing function to the option purchase probability in order to obtain a smooth expected net revenue function. Since the

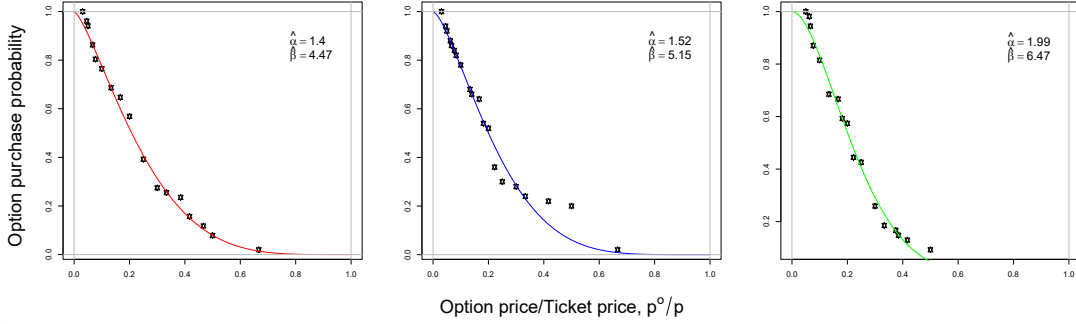


Figure 4.10: Beta probability distribution fits for different bootstrapped samples

beta distribution is defined on the interval $[0, 1]$, which is also the range for the option price to ticket price ratio and the domain of any probability distribution is defined on the interval $[0, 1]$, we consider the beta excess probability distribution as a good choice for this case. This distribution has two shape parameters which we denoted by α, β . To get a good fit, we minimized the mean absolute deviation of the selected beta distribution from the empirical distribution for the option purchase probability $\widehat{g(\cdot)}$. The expression to be minimized with respect to α and β parameters is

$$\alpha^*, \beta^* := \arg \min_{\alpha, \beta} \sum_{i=1}^n \left| \widehat{g(x_i)} - \mathbb{P}\{Y \geq x_i \mid \alpha, \beta\} \right|, \text{ where } Y \sim \text{Beta}(\alpha, \beta).$$

At each bootstrap iteration, optimal shape parameters for beta probability distribution function are updated. In the last two graphs in Figure 4.10, the upper tails of the empirical distributions are underestimated but the distribution shown in the first graph seems like a better fit. To estimate the option purchase probability, instead of using the empirical distribution itself, we fit a excess beta distribution function to this empirical distribution in order to obtain a smooth expected net option revenue function. Thereby we overcame traps into which BFGS sometimes fall on the search for the global maximum of the expected net revenue. Some expected net option revenue functions obtained by fitting beta distributions are shown in Figure 4.11. As we see, the resulting expected net revenue functions are still not concave. Therefore, to make sure that the global maximum is reached, we also extended the initial estimate interval for optimal option price p^o from 0-100 TL to 0-1000 TL. In this way, the search is prevented

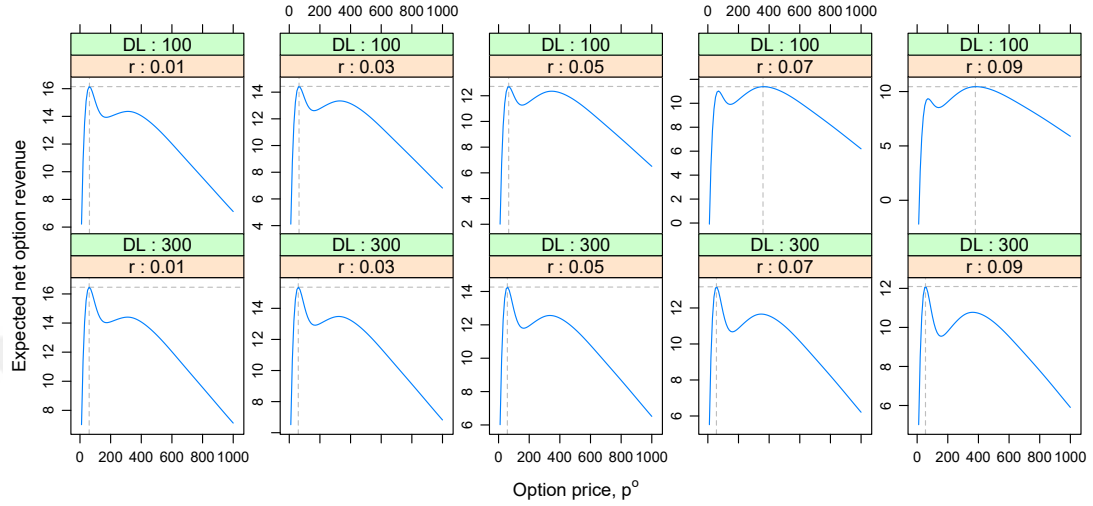


Figure 4.11: After beta distribution is fitted to the option purchasing probability, the expected net option revenue functions are now smooth function of option prices. The maximum net revenues and the corresponding optimal option prices are marked with dashed lines.

from getting stuck in local optima.

The demand of leisure travellers for the airline tickets is more elastic than that of business travellers to the travel costs. Therefore, in the face of increased travel costs, The leisure travellers are more likely to delay their travel plans than the passengers who travel for business. This suggests that, in the absence of the option, the probability of re-ticketing after a missed flight is typically low for the leisure travels and high for the business travels. Moreover, when the missed flight cover is introduced into the market, a higher proportion of the leisure travellers will be willing to buy new tickets after a missed flight than those business travellers. Therefore, the option is likely to create positive cash-flows from option sales as well as the ticket price differences paid by an increased number of leisure travellers who missed their flights. At the same time, some of the business travellers will also take advantage of the option to reduce their costs of missed flights, which they are highly likely to resume even in the absence of the option. This flexibility of the option generates negative cash-flow for the airline from business travellers. Therefore, the option will be profitable if there is

an option price that generates higher positive cash-flows than the magnitude of the negative cash-flows; namely, if the sum of the option fees to be received from leisure and business travels and the additional ticket prices the leisure travellers will pay in case they miss their flights exceeds the full replacement ticket prices that business travellers will reduce by buying a missed flight cover.

In Figure 4.11, for the smaller r values expected net revenue functions attain their maximum at the values less than 200 TL but optimal option price is increasing at the higher values of r . This means that as the probability of a passenger missing her flight increases, optimal option price attains higher values to compensate the negative cash flows generated by the business travels. Since we have a bi-modal expected net option revenue function, we should be careful about both solutions (global and optimal) and have a better knowledge about them because the current solution might switch to the other one. At the smaller mode, the function has a high curvature which makes this solution critical because slight deviations from the true value of optimal option price cause fast declines in expected net option revenue. The difference between the values of the function at both modes is shrinking as the parameter r increases and it becomes insignificant when r is greater than seven percent.

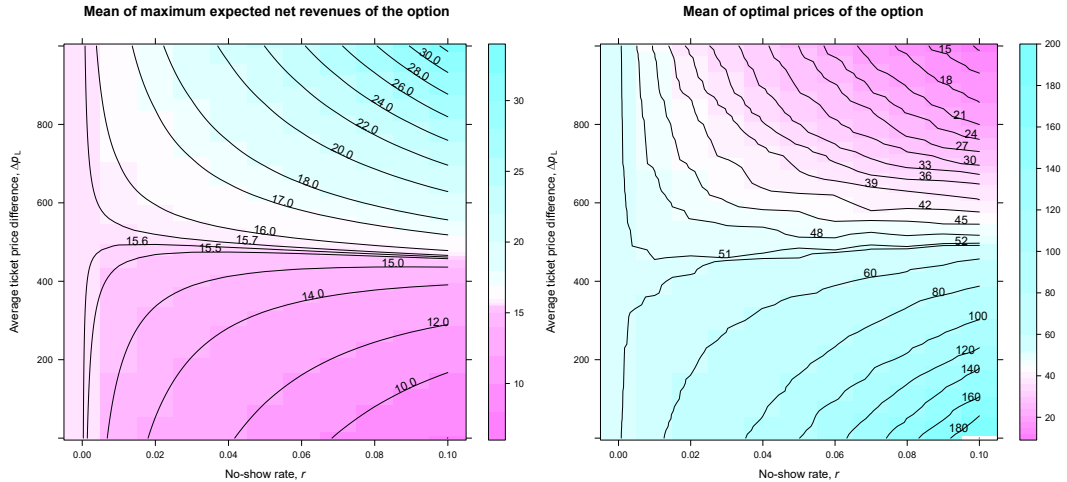


Figure 4.12: After beta distribution fitted, means of maximum expected net option revenues per passenger (left) and optimal option prices (right) over 1,000 bootstrapped samples of market survey data are shown by level and contour graphs.

After the option purchase probability function is smoothed with a beta distribution excess probability function, the means of maximum expected net revenues generated by missed flight cover (left) and the optimal option prices (right) are recalculated and presented with level and contour graphs in Figure 4.12. More detailed results for different r and Δp_L parameters can be found in Table 4.6 and 4.7. Mean maximum expected net option revenue per passenger changes between 8-33 TL whereas mean optimal option price changes between 12-200 TL. Those ranges are obtained as we changed the probability of missing a flight in the interval between 0%-10% and the ticket price differences for the missed and new flights between 0-1000 TL.

Table 4.6: After option purchase probabilities are smoothed with excess beta distribution function, the means (and standard deviations) of maximum expected net option revenues per passenger over 1,000 bootstrapped samples of market survey data

Δp_L	r										
	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
1000	16(2)	17(2)	18(2)	19(3)	20(3)	22(4)	24(5)	26(5)	28(6)	30(7)	33(8)
900	16(2)	16(2)	17(2)	18(3)	19(3)	21(4)	22(4)	24(5)	25(6)	27(7)	29(7)
800	16(2)	16(2)	17(2)	18(2)	18(3)	19(3)	20(4)	21(5)	23(5)	24(6)	25(7)
700	16(2)	16(2)	16(2)	17(2)	17(3)	18(3)	19(4)	19(4)	20(5)	21(6)	22(6)
600	16(2)	16(2)	16(2)	16(2)	17(3)	17(3)	17(3)	18(4)	18(4)	19(5)	19(6)
500	16(2)	16(2)	16(2)	16(2)	16(2)	16(3)	16(3)	16(4)	16(4)	16(5)	17(5)
400	16(2)	15(2)	15(2)	15(2)	15(2)	15(3)	15(3)	14(3)	14(4)	14(4)	14(4)
300	16(2)	15(2)	15(2)	14(2)	14(2)	14(3)	13(3)	13(3)	13(3)	12(4)	12(4)
200	16(2)	15(2)	15(2)	14(2)	13(2)	13(2)	12(3)	12(3)	11(3)	11(3)	10(3)
100	16(2)	15(2)	14(2)	13(2)	13(2)	12(2)	11(2)	11(3)	10(3)	10(3)	9(3)
0	16(2)	15(2)	14(2)	13(2)	12(2)	11(2)	11(2)	10(2)	9(2)	9(2)	8(3)

According to the results in Table 4.6, the effect of no-show rate, r on the expected net revenue is reversed by the change in the value of average ticket price difference for leisure travels Δp_L : if Δp_L is less than 500 TL, then the value of the mean maximum expected net option revenue is a decreasing function of r . In fact, for the greater values of r the expected net revenue becomes negative, and the option is no longer a profitable product. However, as we mentioned earlier, we do not expect the true value of r being greater than ten percent for any airline company. If Δp_L is greater than 500 TL, then the value of the mean of maximum expected net option revenue is an increasing function of r . Essentially, the difference between ticket prices is likely being less than 300 TL

and the probability of missing a flight is likely being between one-to-four percent. In those expected conditions, we can say that the average maximum expected net revenue is greater than 8 TL with 95% confidence. Since the confidence intervals in Table 4.6 do not include zero, it can be said that with the current r and Δp_L ranges the option is expected to provides positive cash flows to the airlines.

Table 4.7: After option purchase probabilities are smoothed with excess beta distribution function, the means (and standard deviations) of optimal option prices over 1,000 bootstrapped samples of market survey data

Δp_L	r										
	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
1000	52(53)	45(50)	38(40)	34(41)	31(41)	27(40)	23(35)	20(36)	17(35)	14(26)	12(27)
900	52(53)	46(51)	40(47)	36(44)	33(42)	30(43)	27(46)	24(42)	21(41)	18(42)	16(41)
800	52(53)	47(52)	43(52)	40(53)	36(50)	33(50)	30(49)	28(50)	26(53)	23(52)	21(52)
700	52(53)	47(52)	45(54)	43(56)	40(58)	39(61)	36(62)	34(63)	31(62)	30(67)	30(73)
600	52(53)	49(55)	47(57)	45(59)	44(63)	44(72)	42(74)	41(77)	41(83)	41(90)	39(92)
500	52(53)	50(56)	49(60)	50(68)	50(73)	49(78)	49(81)	50(91)	50(96)	50(102)	50(107)
400	52(53)	52(58)	53(67)	54(74)	55(80)	56(87)	59(98)	61(104)	66(118)	72(135)	77(147)
300	52(53)	53(61)	55(70)	58(78)	62(90)	66(99)	71(111)	78(126)	86(141)	94(157)	101(170)
200	52(53)	55(63)	59(75)	64(86)	69(96)	76(110)	84(124)	94(140)	103(155)	116(172)	131(191)
100	52(53)	56(65)	62(79)	69(91)	76(104)	87(120)	99(136)	114(155)	127(171)	146(189)	162(204)
0	52(53)	58(68)	66(84)	73(95)	86(112)	100(130)	115(148)	131(163)	153(183)	175(201)	200(218)

Unlike the variance in average maximum expected net revenue, the variance in optimal option price is high. In Table 4.7 the standard deviation is increasing towards the southeast. For the values of Δp_L greater than 800 TL the standard deviation of optimal option price is constantly decreasing with the increase in r values. However, for the values of Δp_L less than or equal to 500 TL, the standard deviation is increasing in the eastward direction.

Chapter 5

Revisiting Option Purchase and Resuming The Missed Flight Models

In our simplified model described in the previous chapter, we modeled the probability of a passenger purchasing the missed flight cover and resuming her journey after a missed flight using aggregate data. We assume that the option purchase probability only depends on the ratio of missed flight cover price and the ticket price. The probability of a passenger resuming a missed flight is determined solely by the motive of her journey. However there may be other factors governing the decision of passengers which we could not measure with the available data. The other factors related to the features of the products and the characteristic traits of passengers can affect the decisions of option purchase and resumption of the travel. Besides the prices of the option and the first ticket, the features of products such as the validity period of the missed flight cover and the price of the second ticket, purchased after the missed flight, can help explaining the passenger behaviors better. Demographic information and the travel habits specific of individuals can also provide insights to the variation in the actions of passengers [8, 9].

Discrete choice models are commonly used in the literature to gather information about customer decisions [10, 11, 12]. Those models are employed to link the discrete outcomes to some observable factors related to both product and the decision maker. Both decisions to purchase the missed flight cover and to resume a missed flight have discrete outcomes. Hence, we believe that discrete choice models are appropriate for further elaboration of those choice situations.

The prominent assumption in discrete choice models is that customers adopt utility maximizing behavior [13]. According to Random Utility Models (RUM), a decision maker would gain a net utility from a specific alternative. The researcher only observes the chosen alternative. The utility generated by this alternative is not observable, but some of its attributes could be associated to decision maker's utility. This part of the utility is called representative utility. Even though some attributes are observable and can be related with the product utility, some factors which govern the decision maker's choice cannot be captured by the researcher. Therefore, RUMs suggest that true utility includes a random term in order to represent those unobserved factors. Discrete choice models differ according to the specification of this random term.

One of the most popular discrete choice models is the logit model [14]. This model assumes that random part of the utility comes from an iid extreme-value distribution. Due to this assumption unobserved factors are modeled as if they are uncorrelated and have the same variance over all alternatives. In some choice situations, this assumption might not be realistic. For example, some unobserved preferences specific to the decision maker might be correlated over some alternatives. Nested logit model relaxes the uncorrelated error terms assumption by grouping similar alternatives in a nest. In this model, the correlation of unobserved factors over alternatives within a nest is allowed whereas the correlation of unobserved factors between nests is prohibited. Probit models allow correlation of unobserved factors by assuming a joint normal distribution for them. However, in some cases error terms may not be normally distributed [15, 16]. Mixed logit is another discrete choice model is free of uncorrelated errors assumption. Unobserved factors are divided into two parts, one of which accounts for all the correlation whereas the other part comes from an iid extreme-value distribution.

This decomposition enables that any discrete choice model can be approximated by mixed logit models [17].

Taste variation may exist between customers. Systematic part of this taste variation can be accommodated by including individual specific factors to the utility function. However, two decision makers who have the same age and income may have different preferences over a particular product. The same customer may even have different preferences for the same choice situation over time. To be able to fully represent the market we need a model that accounts also for random taste variation. Since we have used an aggregate data we could not capture the heterogeneity among passengers. Each passenger may have different sensitivity towards prices and aggregate analysis could not fully represent those differences. With aggregate data we represent a whole distribution with summary statistics. If the population distribution is not uniformly distributed, using an aggregate data may result in misleading inferences for individuals. In the presence of heterogeneity in population, we might be interested not only in average person but also variation between individuals. This will allow us to model passengers behaviors more realistically. Since the mixed logit suggests that model coefficients come from a distribution and vary over individuals rather than being fixed as in logit model, it is an appropriate choice model to address nonuniform actions in the population.

Besides the central tendencies in the population, customers' extreme preferences play an important role in determining the profit generated by a product. Tail areas of the consumer population can be targeted by using product differentiation strategies. Probit and mixed logit models delineate the diversity in the population, however they cannot, by themselves, be used to diagnose or target the individuals who constitute the extreme. Hierarchical Bayes random-effects model is proposed in the literature to draw inferences about specific customers [18]. Characterizing the individuals who are at the extreme also yield more precise estimation of preferences for the population. Therefore, we decided to adopt hierarchical Bayesian approach to model the probabilities of missed flight cover purchase and travel resumption.

5.1 Hierarchical Bayesian Mixed Logit Regression Model

We assume that passenger choices follow a hierarchical Bayesian mixed logit regression model. The utility, that individual h generates from the alternative j in choice set i U_{hij} is defined as

$$U_{hij} = V_{hij} + \varepsilon_{hij},$$

where V_{hij} the representative utility and the i.i.d. extreme-value distributed ε_{hij} , constitute the random component of the true utility. The representative utility is specified as

$$V_{hij} = x'_{hij}\beta_h,$$

where x_{hij} is a p -dimensional vector of attribute levels of alternative i in choice set j , and β_h is a p -dimensional vector of individual h 's part-worths. Since ε_{nij} is distributed extreme value, independent over h , i , and j , the choice probabilities conditional on β_h are defined as

$$p_{hij}(\beta_h) = \frac{\exp(x'_{hij}\beta_h)}{\sum_{t=1}^J \exp(x'_{hit}\beta_h)}. \quad (5.1)$$

The alternative chosen by individual h in choice set i is denoted by a J -dimensional vector, $y_{hi} = (y_{hi1}, \dots, y_{hiJ})$, in which y_{hij} equals to 1 when individual h chooses alternative j and zero otherwise. All chosen alternatives for individual h are denoted by $y'_h = (y'_{h1}, \dots, y'_{hn_h})$ which is a Jn_h -dimensional vector. The model matrix that contains the attribute levels of all alternatives in choice set i evaluated by individual h is represented by a $(J \times p)$ -dimensional matrix, $X_{hi} = (x'_{hi1}, \dots, x'_{hiJ})$. All the choice sets evaluated by individual h are stacked in $X_h = (X_{h1}, \dots, X_{hn_h})$ which is a $(Jn_h \times p)$ -dimensional matrix. The likelihood of individual h 's observed choices, y_h , conditional on β_h , given the choice set design X_h becomes

$$\ell(\beta_h | X_h, y_h) = \prod_{i=1}^{n_h} \prod_{j=1}^J p_{hij}(\beta_h)^{y_{hij}}. \quad (5.2)$$

To model the population, the mixed logit model aggregates individual choice behaviors with a heterogeneity distribution over individual specific part-worths.

The mean vector

$$\beta_h = \Delta z_h + u_h, \quad u_h \sim \mathcal{N}(\mu, \Sigma), \quad (5.3)$$

for heterogeneity distribution of individual specific part-worths are specified as a function of individual specific covariates which include demographic and observed behavior variables. where Δ is a $p \times n_z$ matrix of regression coefficients that relate β_h to the value of z_h , which is a n_z -dimensional vector of covariates. For individual h , z_h elucidates the observed heterogeneity whereas u_h specifies unobserved heterogeneity.

Allenby and Ginter [18] incorporated (5.3) into the random-effect models. In the standard mixed logit model Δ is set to zero and $\beta_h \sim \mathcal{N}(\mu, \Sigma)$. Thus Δ facilitates the identification of individuals whose part-worths are far from the population average. The covariance matrix Σ describes the unobserved heterogeneity in the population.

The unconditional likelihood of individual h is the integral of (5.2) over β_h , which is

$$\begin{aligned} \ell(\Delta, \mu, \Sigma \mid z_h, X_h, y_h) &= \int \ell(\beta_h \mid y_h, X_h) \phi(\beta_h \mid \Delta z_h + \mu, \Sigma) d\beta_h, \\ &= \int \prod_{i=1}^{n_h} \prod_{j=1}^J p_{hij}(\beta_h)^{y_{hij}} \phi(\beta_h \mid \Delta z_h + \mu, \Sigma) d\beta_h, \end{aligned}$$

where ϕ is the normal density function.

We use Bayesian approach for the inferences of part-worths. The prior distribution for the regression coefficient matrix Δ and the random component u_h of individual h 's part-worths are defined in the convenient conditionally conjugate forms

$$\begin{aligned} \text{vec}(\Delta) = \delta &\sim \mathcal{N}(\bar{\delta}, A_\delta^{-1}), \\ \mu &\sim \mathcal{N}(\bar{\mu}, \Sigma \otimes a_\mu^{-1}), \\ \Sigma &\sim \mathcal{IW}(\nu_0, V_0), \end{aligned}$$

where $\bar{\delta}$ is the mean vector, A_δ is the precision matrix for the normally distributed prior of $\text{vec}(\Delta)$, μ follows a p -variate normal distribution with mean $\bar{\mu}$, and

covariance matrix $\Sigma \otimes a_\mu^{-1}$, and Σ follows an inverse Wishart distribution with degrees of freedom ν_0 and $p \times p$ scale matrix V_0 . Then the Directed Acyclic Graphical (DAG) model for the hierarchical Bayesian mixed logit model can be depicted as in Figure 5.1.

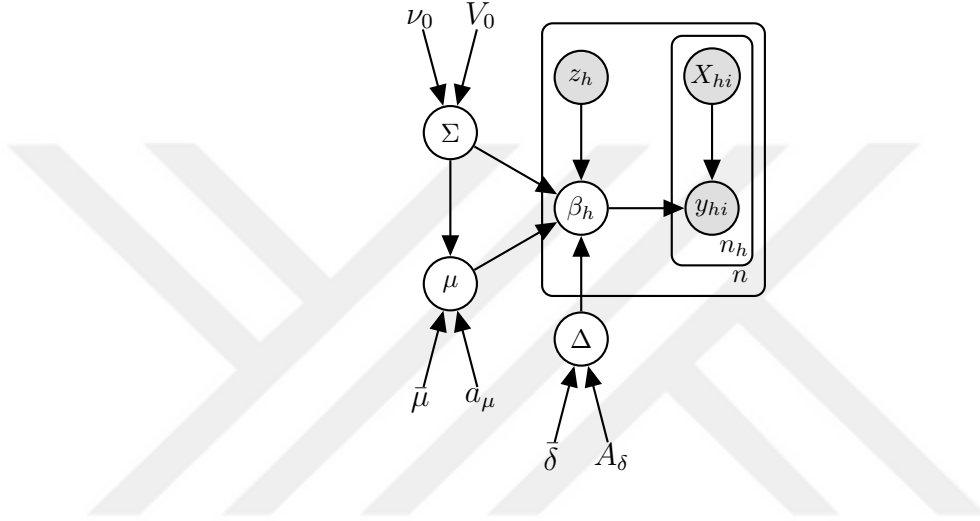


Figure 5.1: DAG model for the mixed logit model coefficients

The joint posterior distribution for our hierarchical Bayesian mixed logit regression model is

$$p(\Sigma, \mu, \Delta, \{\beta_h\}, \{y_h\}) = p(\Sigma) p(\mu | \Sigma) p(\Delta) \prod_{h=1}^n \left(\phi(\beta_h | \Delta z_h + \mu, \Sigma) \prod_{i=1}^{n_h} \prod_{j=1}^J p_{hij}(\beta_h)^{y_{hij}} \right).$$

The full conditional posterior distributions are

$$\begin{aligned} p(\Sigma | \mu, \Delta, \{\beta_h\}) &\propto p(\Sigma | \nu_0, V_0) \phi(\mu | \bar{\mu}, \Sigma \otimes a_\mu^{-1}) \\ &\quad \times \prod_{h=1}^n \phi(\beta_h | \Delta z_h + \mu, \Sigma), \\ p(\mu | \Sigma, \Delta, \{\beta_h\}) &\propto \phi(\mu | \bar{\mu}, \Sigma \otimes a_\mu^{-1}) \prod_{h=1}^n \phi(\beta_h | \Delta z_h + \mu, \Sigma), \\ p(\Delta | \Sigma, \mu, \{\beta_h\}) &\propto \phi(\Delta | \bar{\delta}, A_\delta^{-1}) \prod_{h=1}^n \phi(\beta_h | \Delta z_h + \mu, \Sigma), \end{aligned}$$

$$\begin{aligned}
p(\beta_h \mid \Sigma, \mu, \Delta) &\propto p(\Sigma \mid \nu_0, V_0) \phi(\mu \mid \bar{\mu}, \Sigma \otimes a_\mu^{-1}) \phi(\Delta \mid \bar{\delta}, A_\delta^{-1}) \\
&\times \prod_{h=1}^n \phi(\beta_h \mid \Delta z_h + \mu, \Sigma) \prod_{i=1}^{n_h} \prod_{j=1}^J p_{hij}(\beta_h)^{y_{hij}}. \quad (5.4)
\end{aligned}$$

5.2 Posterior Distributions

We can use the the MCMC chain of Gibbs style defined in [19] by alternating between the draws of individual unit-level parameters and hierarchical parameters:

$$\begin{aligned}
\beta_h &\mid \Delta z_h, \mu, \Sigma, \\
\Delta, \mu, \Sigma &\mid B,
\end{aligned}$$

$$\text{where } B = Z\Delta' + U, \quad B = \begin{bmatrix} \beta'_1 \\ \vdots \\ \beta'_n \end{bmatrix}, \quad U = \begin{bmatrix} u'_1 \\ \vdots \\ u'_n \end{bmatrix}.$$

Given a draw of μ and Σ , a chain can be defined for the draws of β_h . Since there is not a convenient conjugate prior form the conditional posterior of β_h . Therefore, the draws of the posterior distribution in (5.1) is obtained using a Metropolis algorithm. We use a random walk Metropolis algorithm to achieve sampling from the individual unit-level parameters β_h . Random Walk (RW) Metropolis needs an increment density to achieve those draws.

For the parameters of the increment density, mostly maximum likelihood estimators are suggested. Due to the lack of unit-level data, the maximum likelihood estimator (MLE) may not exist (i.e. non-finite MLE.) for a specific unit. Prior distributions can be used to handle this problem. However, this solution requires

recalculation of the moments of RW increment distribution at every MCMC iteration, which makes the algorithm impractical. To overcome the non-finite maximum likelihood estimator problem for individual observations Ross et al. [19] modify unit-level likelihood by multiplying with the pooled likelihood to obtain a pseudo-likelihood which is introduced in (5.2). Let

$$\ell_h^*(\beta) = \ell_h(\beta_h) \bar{\ell}(\beta)^w,$$

where $\ell_h(\beta_h)$ is unit-level, $\bar{\ell}(\beta)$ is the pooled likelihood function, and w is the weight parameter is used to scale the pooled likelihood to prevent it from dominating unit-level likelihood. The weight parameter w is defined as

$$w = \frac{c n_h}{n},$$

where n_h is the number of observations for individual h , n is the total number of observations available in data, and c is the tuning constant which we set it to 0.1.

To obtain a location and scale parameter estimate for increment density of RW, we can maximize the pseudo-likelihood defined in (5.2) and those location and scale parameters can be combined with the prior distribution parameters of individual coefficients. Since we selected a multivariate normal distribution for the prior distribution of individual coefficients, and if normal approximation is applied to pseudo-likelihood, we can easily combine those two to obtain a customized increment density for RW Metropolis algorithm. The proposed location β_h^* and scale Σ_h parameters are

$$\begin{aligned}\beta_h^* &= (H_h + \Sigma^{-1})^{-1} (H_h \hat{\beta}_h + \Sigma^{-1} \bar{\beta}), \\ \Sigma_h &= s^2 (H_h + \Sigma^{-1})^{-1},\end{aligned}$$

where $\bar{\beta}$ is the location parameter obtained by the maximization of pseudo-likelihood in (5.2), $H_h = - \left. \frac{\partial^2 \log(\ell_h^*)}{\partial \beta \partial \beta'} \right|_{\beta = \hat{\beta}_h}$ and s is the scaling parameter for the proposed increment density covariance matrix. We set s to $\frac{2.93}{\sqrt{p}}$ as Roberts and Rosenthal [20] and Ross et al. [19] proposed where p is the dimension of β_h vector.

The draw of hierarchical parameters can be broken down into a succession of

conditional draws:

$$\begin{aligned}\mu, \Sigma &| \Delta z_h, \mu, \Sigma, \\ \Delta &| \mu, \Sigma, B.\end{aligned}$$

The draw of μ and Σ can be achieved using the algorithm to draw from the multivariate regression model (MRM). We employ matrix form for the mixed logit model coefficients to make use of MRM equations. The error components of the mixed logit coefficients are defined as

$$U = B - Z\Delta', \quad \text{where} \quad U = \begin{bmatrix} u'_1 \\ \vdots \\ u'_n \end{bmatrix}.$$

We can decompose those error components as $u_h = 1\mu + \varepsilon_h$ and then they can be expressed in the matrix form as

$$U = \iota\mu' + E, \quad E = \begin{bmatrix} \varepsilon'_1 \\ \vdots \\ \varepsilon'_n \end{bmatrix}, \quad \varepsilon_h \sim \text{iid } \mathcal{N}(0, \Sigma), \quad \iota = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{n \times 1}.$$

Utilizing the notation $\text{etr}(\cdot) \equiv \exp(\text{tr}(\cdot))$ we can write the likelihood functions for E and U as

$$\begin{aligned}p(E | \Sigma) &\propto |\Sigma|^{-n/2} \text{etr} \left\{ -\frac{1}{2} E' E \Sigma^{-1} \right\}, \\ p(U | \mu, \Sigma) &\propto |\Sigma|^{-n/2} \text{etr} \left\{ -\frac{1}{2} (U - \iota\mu')' (U - \iota\mu') \Sigma^{-1} \right\}.\end{aligned}$$

We can decompose the sum of squares term for U , using the least squares projection:

$$\begin{aligned}p(U | \mu, \Sigma) &\propto |\Sigma|^{-n/2} \text{etr} \left\{ -\frac{1}{2} (S_u + (\mu - \hat{\mu})\iota'\iota(\mu - \hat{\mu})') \Sigma^{-1} \right\}, \\ &= |\Sigma|^{-n/2} \text{etr} \left\{ -\frac{1}{2} (S_u + n(\mu - \hat{\mu})(\mu - \hat{\mu})') \Sigma^{-1} \right\}\end{aligned}$$

with $S_u = (U - \iota\hat{\mu}')'(U - \iota\hat{\mu}')$, and $\hat{\mu}' = (\iota'\iota)^{-1}\iota'U = \frac{1}{n} \sum_{h=1}^n u'_h$. Then we can break up the two terms in the exponent:

$$p(U | \mu, \Sigma) \propto |\Sigma|^{-(n-1)/2} \text{etr} \left\{ -\frac{1}{2} S_u \Sigma^{-1} \right\} |\Sigma|^{-1/2} \text{etr} \left\{ -\frac{1}{2} n(\mu - \hat{\mu})(\mu - \hat{\mu})' \Sigma^{-1} \right\}.$$

We know that the joint prior have the form:

$$p(\Sigma, \mu) = p(\Sigma)p(\mu \mid \Sigma)$$

with $\Sigma \sim \mathcal{IW}(\nu_0, V_0)$ and $\mu \mid \Sigma \sim \mathcal{N}(\bar{\mu}, \Sigma \otimes a_\mu^{-1})$. If we combine the likelihood functions for E and U with the joint prior distribution of them, we can obtain a joint posterior which is a product of an inverse Wishart and ‘matrix’ normal kernel.

$$\begin{aligned} p(\Sigma, \mu \mid U) &\propto |\Sigma|^{-(\nu_0+p+1)/2} \text{etr} \left\{ -\frac{1}{2} V_0 \Sigma^{-1} \right\} \\ &\times |\Sigma|^{-1/2} \text{etr} \left\{ -\frac{1}{2} (\mu - \bar{\mu})(\mu - \bar{\mu})' a_\mu \Sigma^{-1} \right\} \\ &\times |\Sigma|^{-n/2} \text{etr} \left\{ -\frac{1}{2} (U - \iota \mu')'(U - \mu') \Sigma^{-1} \right\}. \end{aligned}$$

We can combine the two terms involving μ , using the least squares projection:

$$\begin{aligned} &(\mu - \bar{\mu})(\mu - \bar{\mu})' a_\mu + (U - \iota \mu')'(U - \iota \mu') \\ &= (V - W \mu')'(V - W \mu') \\ &= (V - W \tilde{\mu}')'(V - W \tilde{\mu}') + (\mu - \tilde{\mu}) W' W (\mu - \tilde{\mu})' \end{aligned}$$

with

$$W = \begin{bmatrix} \iota \\ a_\mu^{1/2} \end{bmatrix}, \quad V = \begin{bmatrix} U \\ a_\mu^{1/2} \tilde{\mu}' \end{bmatrix}.$$

The posterior density can now be rewritten as

$$\begin{aligned} p(\Sigma, \mu \mid U) &\propto |\Sigma|^{-(\nu_0+n+p+1)} \text{etr} \left\{ -\frac{1}{2} (V_0 + (V - W \tilde{\mu}')'(V - W \tilde{\mu}') \otimes \Sigma^{-1}) \right\} \\ &\times |\Sigma|^{-1/2} \text{etr} \left\{ -\frac{1}{2} (\mu - \tilde{\mu}) W' W (\mu - \tilde{\mu})' \right\} \end{aligned}$$

with

$$\begin{aligned} (V - W \tilde{\mu}')'(V - W \tilde{\mu}') &= (U - \iota \tilde{\mu}')'(U - \iota \tilde{\mu}') + (\tilde{\mu} - \bar{\mu})(\tilde{\mu} - \bar{\mu})' a_\mu, \\ \tilde{\mu} &= (\iota' \iota + a_\mu)^{-1} (\iota' \iota \hat{\mu} + a_\mu \bar{\mu}) = (n + a_\mu)^{-1} (n \hat{\mu} + a_\mu \bar{\mu}). \end{aligned}$$

Thus, the posterior is in the as form as the conjugate prior: inverse Wishart $\times$ conditional normal.

$$\begin{aligned}\Sigma &| U \sim IW(\nu_0 + n, V_0 + S), \\ \mu &| U, \Sigma \sim \mathcal{N}(\tilde{\mu}, \Sigma \otimes (n + a_\mu)^{-1}), \\ \tilde{\mu} &= (n + a_\mu)^{-1} (n\hat{\mu} + a_\mu\bar{\mu}), \\ S &= (U - \iota\tilde{\mu}')'(U - \iota\tilde{\mu}') + (\tilde{\mu} - \bar{\mu})(\tilde{\mu} - \bar{\mu})' a_\mu.\end{aligned}$$

For the draw of Δ we use the data, which is the last $\{\beta_h\}$ draw and perform one draw from a standard Bayesian regression model. If p is the dimension of the parameter vectors $\{\beta_h\}$, let B, Z be $n \times p$ and $n \times n_z$ arrays, respectively. Let D be the mixed logit model coefficients after extracting the error component mean, μ . This way we have an error component matrix E , with zero mean.

$$D = B - \iota\mu'$$

We can write the model for these observations in the form

$$D = Z\Delta' + E \quad \text{or} \quad D' = \Delta Z' + E'.$$

We will stack those p equations up to see how to standardize:

$$\begin{aligned}\text{vec}(D') &= \text{vec}(\Delta Z') + \text{vec}(E') \\ &= (Z \otimes I_p) \text{vec}(\Delta) + \text{vec}(E'),\end{aligned}\tag{5.5}$$

where $\text{var}(\text{vec}(E')) = I_n \otimes \Sigma$, $\Sigma = R'R$, $\Sigma^{-1} = R^{-1}(R^{-1})'$. Therefore, if we multiply through by $I_n \otimes (R^{-1})'$, this will standardize the error variances to give an identity covariance structure:

$$\begin{aligned}\left(I_n \otimes (R^{-1})'\right) \text{vec}(D') &= \left(Z \otimes (R^{-1})'\right) \text{vec}(\Delta) + z, \\ \text{var}(z) &= I_{n \times p}.\end{aligned}$$

After the standardization, (5.5) can be rewritten as

$$d = X\delta + z,$$

$$\begin{aligned}
\text{where } d &= \left(I_n \otimes (R^{-1})' \right) \text{vec}(D'), \\
X &= \left(Z \otimes (R^{-1})' \right), \\
\delta &= \text{vec}(\Delta).
\end{aligned}$$

Given our prior, $\delta \sim \mathcal{N}(\bar{\delta}, A_{\delta}^{-1})$, we can compute the conditional posterior in the standard normal form

$$\delta \mid d, X, \bar{\delta}, A_{\delta} \sim \mathcal{N} \left((X'X + A_{\delta})^{-1} (X'd + A_{\delta}\bar{\delta}), (X'X + A_{\delta})^{-1} \right).$$

The moments needed for this posterior distribution can be calculated efficiently as follows

$$\begin{aligned}
X'X &= Z'Z \otimes R^{-1} (R^{-1})' = Z'Z \otimes (R'R)^{-1} = Z'Z \otimes \Sigma^{-1}, \\
X'd &= (Z' \otimes R^{-1}) \left(I_n \otimes (R^{-1})' \right) \text{vec}(D') \\
&= \left(Z' \otimes \left(R^{-1} (R^{-1})' \right) \right) \text{vec}(D') \\
&= \left(Z' \otimes (R'R)^{-1} \right) \text{vec}(D') \\
&= \left(Z' \otimes \Sigma^{-1} \right) \text{vec}(D') \\
&= \text{vec}(\Sigma^{-1} D' Z').
\end{aligned}$$

Chapter 6

Choice-Based Conjoint Experiment

Choice-based conjoint experiments are widely used in marketing to learn the preferences of customers over different product offerings. In these experiments, the respondents are asked to select their preferred product or service in a choice set with several alternatives. The design of those alternative sets; namely, the choice sets, is one of the critical points that determines the information carried out by the gathered data. Therefore, to learn the option purchase and the flight resume preferences of airline passengers more realistically, we decided to conduct a choice-based conjoint experiment. To obtain a higher level estimation accuracy of our model parameters, we carefully reviewed the customer choice experiment design literature and decided on an individualized choice experiment. Since each customer has different preferences, an aggregate design which aims at learning the population average may not represent the entire population. Since we choose hierarchical Bayesian mixed logit regression models to estimate the probabilities of the option purchase and the missed flight, the individualized designs are the most suitable experimental designs for this task [21, 22]. The rest of this chapter is organized as follows. In Section 6.1, we present the selected attributes and their levels for the choice experiment. We explain how we design the choice sets in Section 6.2. Section 6.3 describes the implementation of the online web survey.

6.1 Attributes and Levels

Our survey (see Appendices B, C) aims to learn the parameters of both option purchase and the flight resume probability models. Those models are also divided into sub-models for each trip type. Since the trip type variable is constant over a choice set, we cannot learn the effect of this variable with a common model. For the flight resume probability model, one of the variables that govern the resume action is having a missed flight cover. Hence, we need to also learn the effect of that variable, but the purchasing an option is a decision made before the resumption action, and natural selection of the alternatives contained in the choice sets of this experiment should be different flight offerings in order to correctly simulate the real-world situation. Therefore we design separate experiments for each trip type to learn the effect of the option on the flight resumption probability. Consequently we design six different experiments: two for the option purchase, and four for the flight resume models.

The attributes for the option purchase model are option price represented as the percentage of ticket price, validity period and the interaction term of those variables. For the flight resume model, the ticket price and number of days left before the departure time of the second flight are the selected attributes. Since we are interested in the purchase/not purchase and resume/not resume probabilities, and forcing the respondents into making a choice may produce misleading results. Therefore we also include a no-choice dummy to the attribute lists of all the models.

In the beginning of the survey, respondents are asked questions regarding their travel habits. Those are the fraction of their business flights, average minimum and maximum ticket prices for each trip type, and the corresponding early and late purchase times of those average ticket prices. Using the information gathered from those preliminary questions, attribute levels are customized according to each respondent. The levels of the option percentage attribute are 5, 10, 15 and 20 percent, and the validity period levels are 1, 3, 5, 7 and 10. Each respondent's ticket price distribution is adapted to the travel habit information gathered in

the beginning of the survey, option price percentages are customized to each respondent's ticket price distribution. For the flight resume problem, the levels of the ticket price and the departure time attributes are generated by employing an extrapolation scheme where a linear relationship between the ticket prices and purchase times of those tickets are assumed. This way, we make sure that reasonable flight profiles are generated in each choice set. To capture the effect of the missed flight cover to the resume probability, different scenarios with and without the options are generated. Those scenarios contain a trip type and a first ticket price generated equally spaced samples from the respondents' ticket price distribution. The number of questions for each trip type is also customized to the respondent by using her fraction of business flights. In a scenario with the option, the amount of the first ticket price is deduced from the ticket prices of the flight alternatives in choice sets.

6.2 Choice Set Design

For each problem, the choice sets we presented to the respondent consist of four alternatives, one of which is the no choice alternative. We adopt the idea of using Kullback-Leibler (KL) divergence as the design criterion for the individualized choice experiment, since KL divergence requires less computation time compared to the DB metric which is widely used in the efficient design literature [23]. To avoid the computational burden, we decided to use KL divergence. In our experiment, each choice set is optimally picked so as to maximize the expected Kullback-Leibler divergence between subsequent posterior distributions of individualized part-worths. To obtain the posterior distributions of individualized part-worths, we use a sampling scheme in which MCMC sampling is incorporated with importance sampling. Since the convergence of MCMC sampler take more than 10,000 iteration, to employ it for a sequential online web survey is impracticable. Therefore, we decided to run an independent MCMC sampler in the background of the survey to generate each choice set within a more reasonable time.

We learn Δ , μ , and Σ in (5.3) from population using this MCMC sampler. Ideally one would update Δ , μ , and Σ after every response, but MCMC reaches to its stationary distribution after 100,000 iterations and auto-correlations of MCMC samples die out if the lag is 10,000. Therefore, we use the MCMC samples of population parameters Δ , μ , and Σ to calculate the prior mean by customizing it with the covariate information of the respondent. Then for the first question of the respondent this prior distribution is used. For the second and latter questions, the posterior distributions are calculated by the importance sampling using the answers for the previous choice sets.

To calculate the expected KL divergence of a choice set, suppose that respondent h has completed $i - 1$ choice sets, and y_h^{i-1} denotes the selected alternatives in the first $i - 1$ choice sets by this respondent. The respondent's i th choice set is then optimally picked by maximizing

$$\sum_{j=1}^J \pi(y_{hij} | y_h^{i-1}) KL [p(\beta_h | y_h^{i-1}), p(\beta_h | y_h^{i-1}, y_{hij})] \quad (6.1)$$

over all possible choice sets where $p(\beta_h | y_h^{i-1})$ and $p(\beta_h | y_h^{i-1}, y_{hij})$ are updated posteriors of individual part-worths, and the weight for each alternative in choice set i is

$$\pi(y_{hi} | y_h^{i-1}) = \int p_{hij}(\beta_h) p(\beta_h | y_h^{i-1}) d\beta_h.$$

where $p_{hij}(\beta_h)$ is the MNL probability defined in (5.1).

When the definition of KL divergence is applied, (6.1) becomes

$$\sum_{j=1}^J \pi(y_{hij} | y_h^{i-1}) KL [p(\beta_h | y_h^{i-1}), p(\beta_h | y_h^{i-1}, y_{hij})]$$

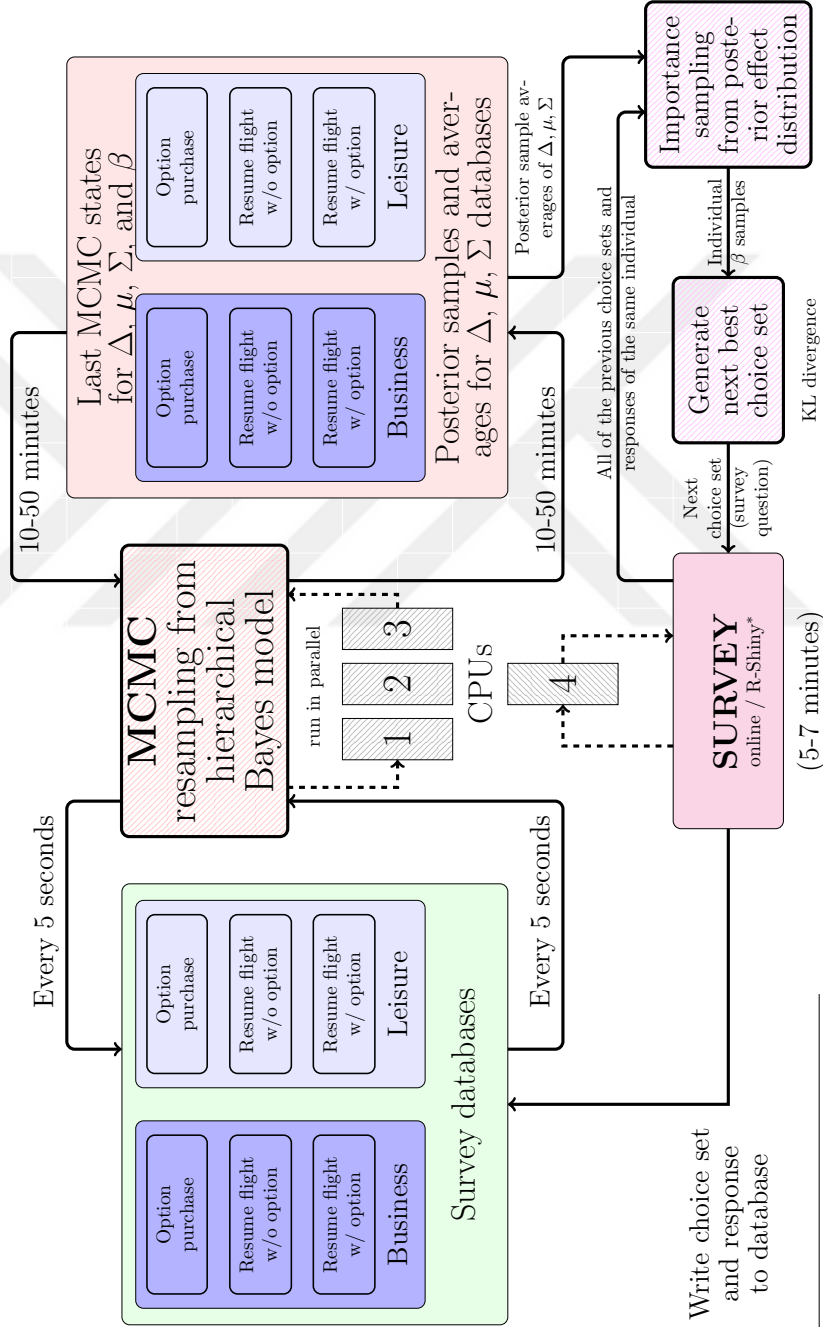
$$\begin{aligned}
&= \sum_{j=1}^J \pi(y_{hij} | y_h^{i-1}) \int p(\beta_h | y_h^{i-1}) \log \frac{p(\beta_h | y_h^{i-1})}{p(\beta_h | y_h^{i-1}, y_{hij})} d\beta_h \\
&= \sum_{j=1}^J \pi(y_{hij} | y_h^{i-1}) \int \log \frac{p(\beta_h | y_h^{i-1}) \pi(y_{hi} | y_h^{i-1})}{p(\beta_h, y_{hij} | y_h^{i-1})} p(\beta_h | y_h^{i-1}) d\beta_h \\
&= \sum_{j=1}^J \pi(y_{hij} | y_h^{i-1}) \int \log \frac{\pi(y_{hi} | y_h^{i-1})}{p_{hij}(\beta_h)} p(\beta_h | y_h^{i-1}) d\beta_h \\
&= \sum_{j=1}^J \pi(y_{hij} | y_h^{i-1}) \left[\log \pi(y_{hij} | y_h^{i-1}) - \int \log p_{hij}(\beta_h) p(\beta_h | y_h^{i-1}) d\beta_h \right].
\end{aligned}$$

KL estimates contain sampling errors. Hence, we used one standard error rule to narrow down the candidates. If the candidate design set contains at least two designs, then we choose the design set that has been asked the least number of times to that individual.

6.3 Implementation

We have six independent MCMC samplers for six problems (see Appendix A) ready to refresh the samples from posteriors when new surveys are added to their respective databases. The program checks every five seconds if a new survey has just been added. If there is at least one idle core, then the program reruns as many MCMC samplers in parallel as the number of idle cores. The rerun priority is given to problems with the least number of updates if the more problems need resampling than available cores. The machine has four cores, but we use at most three of them in parallel at any time and always leave one core for other users of server, including shiny-server which runs independently the online survey application in parallel. The implementation of the online survey is illustrated in Figure 6.1.

The screenshot in Figure 6.2 shows an instance of MCMC resampling process. In the lower left corner, the states of six MCMC samplers are depicted. When activated, each MCMC reads the most recent survey/conjoint data and



* <https://hudson.ie.bilkent.edu.tr:3939/mfs>

Figure 6.1: Implementation of survey

takes 1,000,000 new samples from posterior. Initial 100,000 samples belong to burn-in/warm-up period and are excluded from the study. Every 10,000th of the remaining 1,000,000 samples are held for the use of survey application. Those 100 samples are averaged to get posterior sample mean. Both the samples from posterior distribution and the posterior sample average are stored in MCMC databases. Each survey starts by reading the most up-to-date posterior sample average and use them as the Normal prior for the respondent's effects. The upd column reports the number of times each posterior sample average has been updated (between 104-106 times in this instance). The MCMC samplers ran continuously starting at 02:00 AM on July 5th, 2019, until July 13th, 2019 (except for a two hour disruption due to a minor error). The column running shows the MCMC resampling currently in progress. The column future shows the running time in seconds of last update after it is completed and zero if it is in progress. The mean and sd columns show the average and standard deviations of the running times completed so far. The upper half screen displays the loads on four cores. Two cores were 100% utilized by two MCMC samplings in progress at that moment. Finally, the lower right-hand corner watches for any errors during MCMC runs. In order to intervene promptly after an unexpected crash of MCMC samplers, the server is instructed to send automatic e-mails to the research team as soon as a crash happens.

```

1 [|||||] 100.0% Tasks: 105, 112 thr; 5 running
2 [|||||] 100.0% Load average: 3.98 2.83 1.88
3 [|||||] 100.0% Uptime: 91 days, 20:44:46
4 [|||||] 100.0%
Mem [|||||] 8196/16075MB
Swp [|||||] 538/30516MB

PID USER PRI NI VIRT RES SHR S CPU% MEM% TIME+ Command
19181 shlny 20 0 376M 195M 12904 R 100. 1.2 4:11.05 /usr/lib/R/bin/exec/R --slave --no-restore --vanilla --file=globalMCMC_async4_FINAL.R
19180 shlny 20 0 362M 187M 12904 R 100. 1.2 4:10.40 /usr/lib/R/bin/exec/R --slave --no-restore --vanilla --file=globalMCMC_async4_FINAL.R
19179 shlny 20 0 374M 198M 12904 R 100. 1.2 4:11.78 /usr/lib/R/bin/exec/R --slave --no-restore --vanilla --file=globalMCMC_async4_FINAL.R
12396 shlny 20 0 339M 161M 16836 S 0.5 1.0 31:58.00 /usr/lib/R/bin/exec/R --slave --no-restore --vanilla --file=globalMCMC_async4_FINAL.R
12397 shlny 20 0 14604 1536 1288 S 0.0 0.0 0:00.00 bash -c Rscript --vanilla globalMCMC_async4_FINAL.R 1> data/globalMCMC_async4_FINAL.out 2>
12398 shlny 20 0 7160 1844 1744 S 0.0 0.0 0:00.00 tee data/globalMCMC_async4_FINAL.err
12399 shlny 20 0 10184 1460 1324 S 0.0 0.0 0:00.00 sed 1s/^/From: Hudson Bilkent<mailto:sdayanik@bilkent.edu.tr;beyza.cellk@bilkent.edu.tr\nSubjec
12400 shlny 20 0 7160 1848 1744 S 0.0 0.0 0:00.00 tee -a /dev/fd/63 /dev/fd/62
12401 shlny 20 0 14608 1732 1480 S 0.0 0.0 0:00.00 bash -c Rscript --vanilla globalMCMC_async4_FINAL.R 1> data/globalMCMC_async4_FINAL.out 2>
12402 shlny 20 0 14608 1668 1416 S 0.0 0.0 0:00.00 bash -c Rscript --vanilla globalMCMC_async4_FINAL.R 1> data/globalMCMC_async4_FINAL.out 2>
12403 shlny 20 0 18004 3488 3032 S 0.0 0.0 0:00.00 smtp sdayanik@bilkent.edu.tr

F1Help F2Setup F3Search F4Filter F5Free F6PortB F7Vlce F8Vlce F9Kill F10Quit

Every 2.0s: ./watch_debug.sh Wed Jul 17 07:52:29 2019 Every 2.0s: ./watch_err.sh Wed Jul 17 07:52:29 2019

job ncomp R keep warnup upd running future mean sd
optionLeis 1 1100000 10000 100000 106 FALSE 3720.87 2384.83 903.79
resumeLeis 1 1100000 10000 100000 106 FALSE 2413.16 1541.55 583.48
optionBus 1 1100000 10000 100000 104 TRUE 0.00 1856.77 754.31
resumeBus 1 1100000 10000 100000 105 TRUE 0.00 1240.93 510.48
resumeLeisOpt 1 1100000 10000 100000 105 TRUE 0.00 1619.36 590.85
resumeBusOpt 1 1100000 10000 100000 105 FALSE 1836.41 1137.98 462.53

Attaching packages
tidyverse 1.2.1 —
✓ ggplot2 3.1.1 ✓ purrr 0.3.2
✓ tibble 2.1.1 ✓ dplyr 0.8.1
✓ tidyr 0.8.3 ✓ stringr 1.4.0
✓ readr 1.3.1 ✓ forcats 0.4.0

Conflicts:
tidyverse
* tidy::extract() masks magrittr::extract()
* dplyr::filter() masks stats::filter()
* dplyr::lag() masks stats::lag()
* purrr::set_names() masks magrittr::set_names()
Parsed with column specification:
cols(
  name = col_character(),
  dbname_survey = col_character(),
  dbpath_survey = col_character(),

```

Figure 6.2: An instance of MCMC resampling process

Chapter 7

Numerical Analysis

We incorporated into the model defined in Chapter 3 the new option purchase and the missed flight resume models, introduced in Chapter 5, and implemented the survey described in 6. In this chapter, we will analyse the conjoint data collected with the online survey between May 28 and July 13, 2019. We initially collected data from 30 respondents on May 28 and 29. We studied convergence and autocorrelation of the MCMC. Due to the long convergence time, we decided to do importance sampling, rather than MCMC for individual surveys, rescaled both option price, interaction term with validity, ticket price and demographic variables so that all have the same unit scale. The attribute level scales for option purchase and missed flight resume probabilities are shown in Table 7.1.

We created a reference respondent demographics. During hierarchical Bayesian model treatment and importance sampling, in the calculation of effect means described in (5.3), raw $\tilde{z}_h$ of the respondent will be centered at the reference respondent; namely use $z_h = \tilde{z}_h - \bar{z}$. This way, (5.3) is the distribution of β for the reference respondent. We created also the scales for demographic variables. We divided z components by corresponding scale values to bring all demographic variables to roughly the same scale. This allows us to use the same hyperparameters for the normal priors of Δ . See Appendix D for the reference respondent and demographic variables scale values.

Table 7.1: Attribute level scales of option purchase and flight resume probability functions

Variable	Value	Scaled
Option price percentage (pp^o)	5%, 10%, 15%, 20%	5, 10, 15, 20
Validiy period (v^o)	1, 3, 5, 7, 10	1, 3, 5, 7, 10
Interaction ($pp^o : v^o$)	5%, 15%, 21%, $\dots$ %	0.5, 1.5, 2.1, $\dots$
Ticket price (P)	100, 150, 300, 550, $\dots$	1, 1.5, 3, 5.5, $\dots$
Departure Time	1, 3, 5, 7, 10	1, 3, 5, 7, 10

Within the modified model, the effect of both price percentage and validity period of the missed flight cover are examined with a market simulation. First, we estimate the expected behaviors of passengers through the random effects distributions used in mixed logit models for the option purchase and the missed flight resumption, then we use those estimations to calculate the expected net revenue generated by a typical passenger when the missed flight cover is present. Different missed flight cover designs are evaluated under probable passenger no-show rates.

The rest of this chapter is organized as follows. In Section 7.1, we explore the demographic information gathered using the online web survey and the heterogeneity between passengers. In Section 7.2, we present a generative process that explains how we estimate the random variables included in our graphical model. Lastly, in Section 7.3, we present our results for the expected net revenue generated by the missed flight cover.

7.1 Exploring Data

The market is simulated based on the data gathered with the individualized Bayesian choice-based conjoint experiment. In this experiment, respondents are first asked to provide some self-explicated information including both their demographic information and travel habits, which we denote as covariate variables in our model. Those covariate variables are described in Table 7.2.

Table 7.2: Descriptions of covariates

Covariates	Levels
Gender	“Female” = 1, “Male” = 2
Age	“<25” = 1, “25-35” = 2, “36-45” = 3, “46-60” = 4, “60” = 5)
Income	“<4K” = 1, “4-6K” = 2, “6-10K” = 3, “10-15K” = 4, “15-20K” = 5, “>20K” = 6
Leis Freq	(Fraction of leisure flights) continuous between 0 and 1
Leis Ticket	continuous
Price Mean	
Leis Ticket	(Difference between the maximum and minimum average
Price Span	ticket price for a leisure travel) continuous
Leis Purchase	continuous
Time Mean	
Leis Purchase	(Difference between the earliest and latest a leisure travel is
Time Span	booked) continuous
Bus Freq	(Fraction of business flights) continuous between 0 and 1
Bus Ticket	continuous
Price Mean	
Bus Ticket	(Difference between the maximum and minimum average
Price Span	ticket price for a business travel) continuous
Bus Purchase	continuous
Time Mean	
Bus Purchase	(Difference between the earliest and latest a business travel is
Time Span	booked) continuous

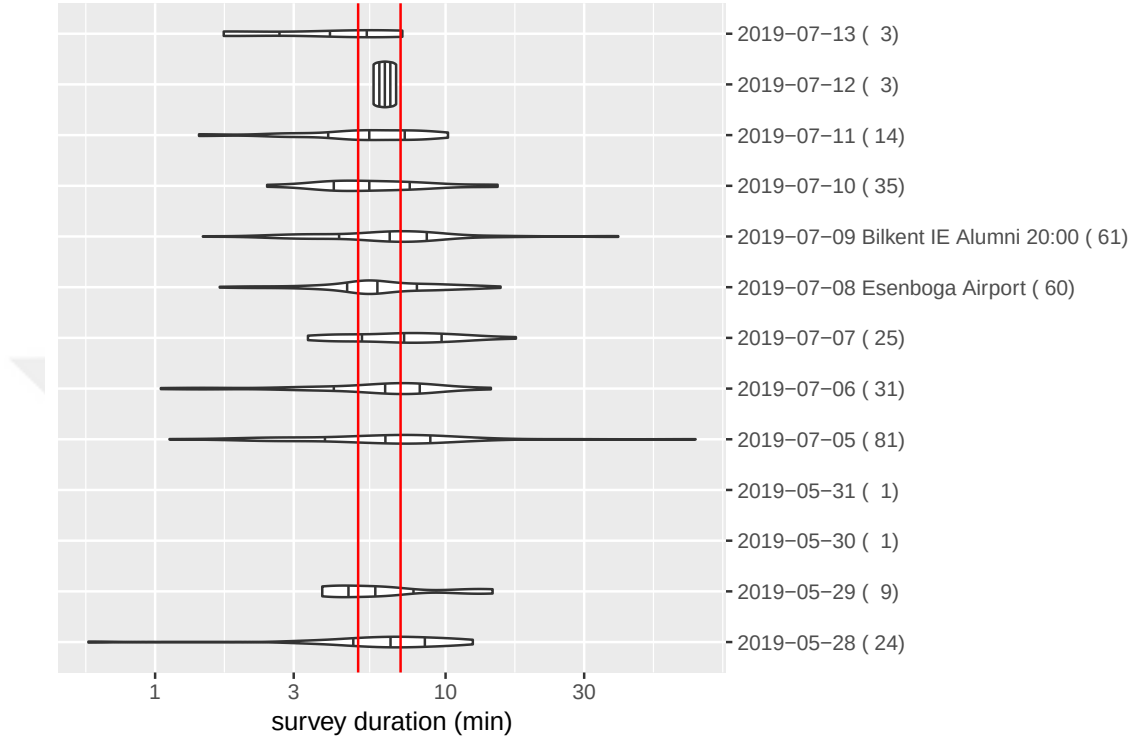


Figure 7.1: Breakdown of the number of respondents into the dates that survey stays online and boxplots of survey durations. Two vertical lines were mark 5 and 7 minutes.

Randomly selected 348 people are requested to fill out our online web survey. The breakdown of the number of respondents into some the dates that the survey stays online with some of their contact channels are illustrated in Figure 7.1.

For each gender, the respondents income distribution is shown in Table 7.3. The division of the total number of respondents into genders seems balanced. Therefore, we do not observe any gender bias in our data. According to that table, lower income intervals contain more female respondents than male respondents whereas the upper intervals contain more male respondents than females. The mass of income marginal distribution is accumulated on the interval of 6 – 10K or smaller intervals. The mass of conditional distribution of income given the respondent is male however, has a fatter tail on the right than the marginal distribution. On the other hand, the conditional distribution of income with respect to female respondents is more right-skewed than both marginal and the

conditional distribution on male.

Table 7.3: Gender versus income

	<4K	4-6K	6-10K	10-15K	15-20K	>20K	Sum
Female	46	53	44	10	4	4	161
Male	34	33	57	32	8	23	187
Sum	80	86	101	42	12	27	348

Table 7.4: Age versus income

	<4K	4-6K	6-10K	10-15K	15-20K	>20K	Sum
<25	23	9	4	1	0	0	37
25-35	56	68	81	26	4	9	244
36-45	1	8	11	13	5	10	48
46-60	0	1	5	2	3	7	18
>60	0	0	0	0	0	1	1
Sum	80	86	101	42	12	27	348

The conditional income with respect to age is given in Table 7.4. Younger respondents have lower income and the marginal distribution of age is concentrated on the 25 – 35 interval.

Table 7.5: Statistics for the covariates of passengers who use some of their flights for business purposes (TP = Ticket Price, PT = Purchase Time)

Freq	TP Mean	TP Span	PT Mean	PT Span
0.4026	355.1	338.3	11.63	14.62

Table 7.6: Statistics for the covariates of passengers who use some of their flights for leisure purposes (TP = Ticket Price, PT = Purchase Time)

Freq	TP Mean	TP Span	PT Mean	PT Span
0.7029	277.2	290	22.89	27.39

The travel habits of respondents are presented in Tables 7.5 and 7.6. We can infer from those tables that when passengers travel for business, their price sensitivity is lower and they book their flights later than when they travel for leisure. In the light of those summary statistics of data, we can once again say that passengers show distinct behaviors according to their trip motives. The data gathered using online survey shows consistent passenger patterns with the data presented in Chapter 4.

7.2 Generative Process

Each passenger has different frequency of business (H) and leisure travels (L), and we use those frequencies to estimate the distribution of T ,

$$\mathbb{P}(T_h) = \begin{cases} \pi_{hH}, & \text{if } T = H, \quad \text{or} \\ \pi_{hL}, & \text{if } T = L, \end{cases}$$

where π_{hL} and π_{hH} represents the frequency of leisure and business travels for passenger h , respectively, and $\pi_{hL} + \pi_{hH} = 1$.

We also customize the ticket price since each passenger has different travel habits and different sensitivity towards to the price. From the aggregate data, we also observe that passengers show different behavior patterns when they travel for business and leisure. We decided to use the mean ticket prices for each trip type which is available in our covariate data for each passenger. However, besides the heterogeneity between passengers and their trip motive, some randomness still exists in ticket prices. To represent this randomness, we assume that the ticket

price for each trip type follows a normal distribution whose parameters are the mean ticket price, and the ticket price span values available in our individual-level data. Therefore, the ticket price of passenger h when she travels for trip type T is

$$P_{1,hT} \sim \mathcal{N}(\mu_{hT}, \sigma_{hT}), \quad (7.1)$$

where $P_{1,hT}$ denotes the first ticket price for passenger h when she travels for trip type T , μ_{hT} is the mean ticket price for the corresponding passenger, when she travels for trip type T , and σ_{hT} equals to one sixth of ticket price span of trip type T for passenger h .

The probability of option purchase is defined with a mixed logit model which is detailed in Chapter 5. The part-worths of this model are determined as no purchase, the option price represented as the percentage of ticket price, pp^o , validity period, v^o and the interaction term between the last two variables. To calculate the option purchase probability, we compare the utilities of different option profiles with no purchase part-worth utility. Hence, the option purchase probability is

$$G_{hT}(pp^o, v^o) = \int \frac{1}{1 + \exp(V_{hTj}(\beta_{hT}^o) - \beta_{hT1}^o)} p(\beta_{hT}^o) d\beta_{hT}^o,$$

$$V_{hTj}(\beta_{hT}^o) = x_{hj}^{o'} \beta_{hT}^o \quad \forall j,$$

where β_{hT1}^o is no-purchase part-worth utility and x_{hj}^o is a vector containing the attribute information of option profile j . To find the expectation of $G_T(\beta_{hT}^o, pp^o, v^o)$ over β_{hT}^o , we draw samples from its posterior distribution, whose derivation is given in Section 5.2, by using MCMC sampler.

The probability of the flight resume is described with a separate mixed logit model. The part-worths for this model are no-resume, the second ticket price or the price difference between the first and the second tickets if the passenger has the option and departure time of the second flight. To calculate those probabilities we need the probability density functions of both second ticket price, $P_{2,hT}$ and ticket price difference, ΔP_{hT} for each passenger and the trip type. Recall that ΔP_{hT} is a function of P_2 and P_1 and it is defined as

$$\Delta P_{hT} := \max(P_{2,hT} - P_{1,hT}, 0). \quad (7.2)$$

Hence, if we learn the value of $P_{2,hT}$ from available data we can easily find the ΔP_{hT} using (7.2).

Since the ticket price is highly correlated with the booking time of the flight, we also decided to incorporate purchase time of the ticket, defined as the number of days until the flight departure date, in the estimation process of the second ticket price. We assume a negative linear relationships between the mean ticket price and mean purchase time of the ticket. Since the minimum and maximum values for average ticket prices and the latest and earliest values for average purchase times are available in the data, we can easily find the slope of that line which approximates the change of average ticket price per day. We denote m_T as the change of average ticket price per day for trip type T and the sampling distribution of that random variable can be defined as

$$\begin{aligned}\tilde{m}_T &\sim [m_{1T}, \dots, m_{hT} \dots, m_{NT}], \\ m_{hT} &= \frac{P_{hT}^l - P_{hT}^e}{l_{hT} - e_{hT}},\end{aligned}\tag{7.3}$$

where m_{hT} represents the change of average ticket price per day given the spans of mean ticket price $P_{hT}^l - P_{hT}^e$ and purchase time $e_{hT} - l_{hT}$ of respondent h and N denotes the number of respondents in our data. P_{hT}^e and P_{hT}^l denote the average minimum and maximum ticket prices for respondent h when she travels for a T type trip, and e_{hT}, l_{hT} denote the corresponding purchase times of those tickets.

To generate the second ticket price, we also need the departure time of the second ticket defined as the number of days from this ticket is booked till the departure. Since we try to calculate the expected net revenue generated by a typical passenger when the missed flight cover is present, we use the validity period of the option as the corresponding departure time of the flight. This way the expected revenue generated by the case without the option is restricted with the time of the option validity period. Thereby, a more reasonable comparison between the expected revenues generated by the cases where the missed flight cover exists and does not exist is made. By using the draw from the sampling distribution of $\tilde{m}_T$ and the mean and span of both ticket price and purchase time covariate information from respondent h we can calculate the second ticket price

for that respondent as

$$P_{1,hT} := P_{hT}^e + m_{hT} (v^o - e_{hT}),$$

where P_{hT}^e denotes the average minimum ticket price indicated by the respondent h for trip type, T and e_{hT} is the corresponding purchasing time for that ticket. After defining the second ticket price and the price difference between the first and ticket prices, we can introduce the probability of a passenger to resume her journey. Let us firstly elucidate the case without the option:

$$H_{hT}(P_{2,hT}, v^o) = \int \frac{1}{1 + \exp(V_{hTj}(\beta_{hT}^r) - \beta_{hT1}^r)} p(\beta_{hT}^r) d\beta_{hT}^r,$$

$$V_{hTj}(\beta_{hT}^r) = x_{hj}^r{}' \beta_{hT}^r \quad \forall j,$$

where β_{hT1}^r is no resume part-worth utility, and x_{hj}^r is a vector containing the attribute information of flight profile j . To find the expectation of $H_T(\beta_{hT}^r, P_{2,hT}, v^o)$ over β_{hT}^r , we draw samples from its posterior distribution by using MCMC sampler. Then the case with the option is defined as

$$H_{hT}(\Delta P_{hT}, v^o) = \int \frac{1}{1 + \exp(V_{hTj}(\beta_{hT}^{ro}) - \beta_{hT1}^{ro})} p(\beta_{hT}^{ro}) d\beta_{hT}^{ro},$$

$$V_{hTj}(\beta_{hT}^{ro}) = x_{hj}^r{}' \beta_{hT}^{ro} \quad \forall j,$$

where β_{hT1}^{ro} is no resume part-worth utility with the option and x_{hj}^r is a vector containing the attribute information of flight profile j . To find the expectation of $H_T(\beta_{hT}^{ro}, \Delta P_{hT}, v^o)$ over β_{hT}^{ro} , we draw samples from its posterior distribution by using MCMC sampler.

Finally, we need to find the expected net revenue to calculate the terms of the (3.6). The first term can be easily calculated as follows

$$\mathbb{P}_h(B) p^o = \sum_{i \in L, H} \pi_{hi} \mu_{hi} G_{hi}(pp^o, v^o). \quad (7.4)$$

For the second term, we need to take expectation over random variables included in the expression, which are $P_{2,hT}$ and ΔP_{hT} . Since we know the sampling distribution of m_{hT} , we can take samples of the distribution given in (7.2) with replacement so that we obtain a new $P_{2,hT}$ observation for each sampling iteration. For the random variable ΔP_{hT} , sampling from the distribution of the first ticket

price, $P1_{hT}$ is also needed. Since we already assume that $P1_{hT}$ follows a normal distribution with a known mean and variance given in the expression (7.2), we can easily draw samples from this normal distribution for each sampling iteration. If we denote a sampling iteration with s , the second term in the equation (3.6) takes the form

$$\begin{aligned} & \mathbb{E}1_{L \cap R}(\Delta P + P1_{B^c}) - \mathbb{E}^*1_{L \cap R}(\Delta P + P) \\ &= \frac{r}{S} \sum_{i \in L, H} \pi_{hi} \sum_{s=1}^S \left[\Delta P_{hi}^s H_{hi}(\Delta P_{hi}^s, v^o) - P2_{hi}^s H_{hi}(P2_{hi}^s, v^o) \right] G_{hi}(pp^o, v^o), \end{aligned} \quad (7.5)$$

where S is the number of sampling iteration.

Now, we can bring those terms expressed in (7.2) and (7.2) together to rewrite the expected net revenue for a typical passenger h as

$$\begin{aligned} & \mathbb{E} [P + p^o 1_B + 1_{L \cap R}(\Delta P + P1_{B^c})] - \mathbb{E}^* [P + 1_{L \cap R}(\Delta P + P)] \\ &= \sum_{i \in L, H} \pi_{hi} G_{hi}(pp^o, v^o) \left[\mu_{hi} + \frac{r}{S} \sum_{s=1}^S (\Delta P_{hi}^s H_{hi}(\Delta P_{hi}^s, v^o) \right. \\ & \quad \left. - P2_{hi}^s H_{hi}(P2_{hi}^s, v^o)) \right]. \end{aligned} \quad (7.6)$$

7.3 Results For Expected Net Revenue

The expected net revenue (7.6) is evaluated for each passenger available in our data by using a discrete grid where the option price percentage, pp^o changes between 5 – 20%, validity period of the option, v^o between 1 – 10 days, and the passenger no-show rate, r takes values between 0.01 – 0.1, respectively. The results for average passenger are demonstrated with level and contour plots in Figure 7.2 in which there are 11 different scenarios generated by the passenger no-show rate, r .

According to Figure 7.2, all missed flight cover designs represented on the discrete grid generates a positive profit value. For the values of option price percentage, pp^o less than or equal to 5%, the expected net revenue generated by

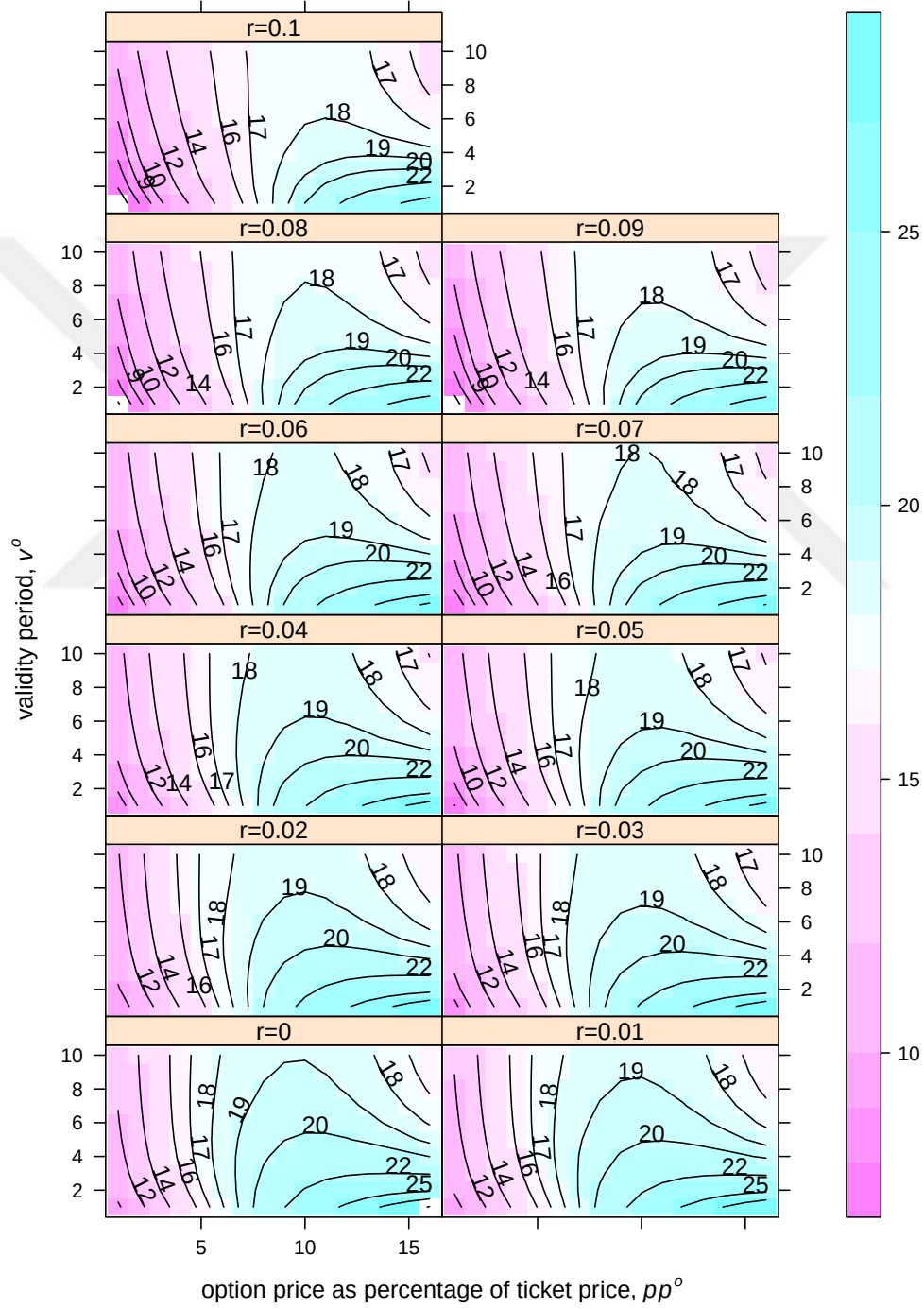


Figure 7.2: Expected net revenue generated by a typical passenger when the missed flight cover is present for different passenger no-show rates

a typical passenger is not affected by the validity period. However, for the values of option price percentage greater than 5%, as the validity period lasts longer, the expected net revenue per passenger is decreasing. The reason for that may be the decrease in the difference between ticket prices, as the validity period lasts longer, the price of the second ticket decreases and the airline loses money from the passengers who are likely to continue their journey, even in the absence of the missed flight cover.

Table 7.7: Optimal designs and corresponding expected net revenue values for every passenger no-show rates

option price percentage	validity period	no-show rate	expected net revenue
20	1	0	29.1
20	1	0.01	28.81
20	1	0.02	28.53
20	1	0.03	28.24
20	1	0.04	27.97
20	1	0.05	27.66
20	1	0.06	27.39
20	1	0.07	27.08
20	1	0.08	26.8
20	1	0.09	26.53
20	1	0.1	26.24

The maximum expected net revenue for a typical passenger is selected over different option design profiles for each value of passenger no-show rate, r on the grid, and the results are presented in Table 7.7. For each passenger no-show rate values, the same missed flight cover design maximizes the expected net revenue. The option which has the highest price percentage and the shortest validity period generates the maximum expected net revenue per passenger. However, the grid search may be stopped at the edge of search space. Therefore, the missed flight cover option designs with higher option price percentage and shorter validity

period can be investigated.



Chapter 8

Conclusion

In this thesis, we study optimal price and validity period of missed flight cover. This option allows a passenger to use a missed flight fare towards the purchase of a future airline ticket. Our aim was to maximize the expected net revenue generated by a typical passenger when the missed flight cover is present. To find the expected net revenue, we firstly model the relationships between all of the random variables that govern the behaviors of passengers. The probabilistic actions of airline passengers are described with a probabilistic graphical model and using that model, joint probability distribution of the random variables are expressed. By means of that graphical model without making any unrealistic assumption like statistical independence between variables, we could calculate the expected net revenue with simulation.

To check the feasibility of our model we conduct a preliminary study using a small data set. From that data set, we realized that passengers may show different behavior patterns when they travel for business and leisure. Then we used Bayesian expectation maximization algorithm to fit a mixture of Normals distribution to ticket prices of for each trip type. For the option purchase probability, firstly we used the empirical distribution. However, we faced some difficulties during the maximization of the expected net revenue function over the option price due to the non-smoothness of the objective function. To circumvent this problem,

we fit a beta distribution to the option purchase probability by minimizing the mean absolute deviation. Since this data set is not gathered taking our graphical model into account, we could not learn the flight resume probability from that data. Instead, we assumed that business travellers almost always resume their journey, whereas the leisure travellers only resume when the option is present. To be able to learn how much uncertainty lies on our results, we used bootstrap algorithm. After observing the sampling means of and standard deviations of both expected net revenues and the option price, we discovered that even though the expected net revenue mostly attains positive values, the standard deviation of optimal option price is large.

To learn the option purchase and the missed flight cover probabilities more realistically, we unified discrete choice and probabilistic graphical models and used two separate hierarchical Bayesian mixed logit regression model. To estimate those model parameters we conducted an online individualized choice-based conjoint experiment which draws samples from the posterior distribution to maximize the KL divergence between the subsequent posterior distributions individualized model parameters. In both data generation and estimation, we took the population heterogeneity into account to obtain precise estimates of individualized part-worths.

Lastly, we found the most profitable option design. In the face of high operating costs and competition in Turkish air travel market, airline companies cannot neglect a new contract capable of producing additional ancillary revenues. With the contract in place, passengers who miss their flights due to minor excuses will be able to resume their journey at much lower costs. The missed flight cover will also ease the jobs of airline agents who have to deal at times with often distressed passengers because of a missed flight.

Bibliography

- [1] F. Alamdari, “Airline in-flight entertainment: the passengers’ perspective,” Journal of Air Transport Management, vol. 5, no. 4, pp. 203–209, 1999.
- [2] J. C. Martín, C. Román, and R. Espino, “Willingness to pay for airline service quality,” Transport Reviews, vol. 28, no. 2, pp. 199–217, 2008.
- [3] K. Mason, “The value and usage of ticket flexibility for short haul business travellers,” Journal of Air Transport Management, vol. 12, no. 2, pp. 92–97, 2006.
- [4] C. M. Bishop, Pattern Recognition and Machine Learning. Springer, 2006.
- [5] T. Minka, “Expectation-Maximization as lower bound maximization,” Tutorial published on the web at <http://www-white.media.mit.edu/tpminka/papers/em.html>, vol. 7, p. 2, 1998.
- [6] A. Corduneanu and C. M. Bishop, “Variational Bayesian model selection for mixture distributions,” in Artificial intelligence and Statistics, vol. 2001, pp. 27–34, Morgan Kaufmann Waltham, MA, 2001.
- [7] S. N. Wood, Core Statistics, vol. 6. Cambridge University Press, 2015.
- [8] G. M. Allenby, P. E. Rossi, and R. E. McCulloch, “Hierarchical Bayes Models: A Practitioners Guide. SSRN Scholarly Paper ID 655541,” Social Science Research Network, Rochester, NY, 2005.
- [9] B. Orme and J. Howell, “Application of covariates within Sawtooth Software’s CBC/HB program: theory and practical example,” Sawtooth Software Research Paper Series, 2009.

- [10] K. E. Train, Discrete Choice Methods with Simulation. Cambridge university press, 2009.
- [11] D. A. Hensher and W. H. Greene, “The mixed logit model: the state of practice,” Transportation, vol. 30, no. 2, pp. 133–176, 2003.
- [12] D. A. Hensher, J. M. Rose, and W. H. Greene, Applied Choice Analysis: A Primer. Cambridge University Press, 2005.
- [13] J. Marschak, “Binary-choice constraints and random utility indicators (1960),” in Economic Information, Decision, and Prediction, pp. 218–239, Springer, 1974.
- [14] D. McFadden et al., “Conditional logit analysis of qualitative choice behavior,” 1973.
- [15] J. A. Hausman and D. A. Wise, “A conditional probit model for qualitative choice: discrete decisions recognizing interdependence and heterogenous preferences,” 1976.
- [16] C. Daganzo, Multinomial Probit: the Theory and its Application to Demand forecasting. Elsevier, 2014.
- [17] D. McFadden and K. Train, “Mixed MNL models for discrete response,” Journal of applied Econometrics, vol. 15, no. 5, pp. 447–470, 2000.
- [18] G. M. Allenby and J. L. Ginter, “Using extremes to design products and segment markets,” Journal of Marketing Research, vol. 32, no. 4, pp. 392–403, 1995.
- [19] P. E. Rossi, G. M. Allenby, and R. McCulloch, Bayesian Statistics and Marketing. John Wiley & Sons, 2012.
- [20] G. O. Roberts, J. S. Rosenthal, et al., “Optimal scaling for various metropolis-hastings algorithms,” Statistical science, vol. 16, no. 4, pp. 351–367, 2001.
- [21] J. Yu, P. Goos, and M. Vandebroek, “Individually adapted sequential bayesian conjoint-choice designs in the presence of consumer heterogeneity,”

International Journal of Research in Marketing, vol. 28, no. 4, pp. 378–388, 2011.

[22] M. Crabbe and M. Vandebroek, “Improving the efficiency of individualized designs for the mixed logit choice model by including covariates,” Computational Statistics & Data Analysis, vol. 56, no. 6, pp. 2059–2072, 2012.

[23] M. Crabbe, D. Akinc, and M. Vandebroek, “Fast algorithms to generate individualized designs for the mixed logit choice model,” Transportation Research Part B: Methodological, vol. 60, pp. 1–15, 2014.

Appendix A

Problem Definitions for Different MCMC samplers

Table A.1: Descriptions of six problems modeled in the survey

Problem	Description
optionLeis	Respondent is asked to choose one of the cover options among several alternatives including no choice for her leisure travels
optionBus	Respondent is asked to choose one of the cover options among several alternatives including no choice for her business travels
resumeLeis	Respondent is asked to choose one of the flights among several alternatives including no choice to resume her leisure travel
resumeBus	Respondent is asked to choose one of the flights among several alternatives including no choice to resume her business travel
resumeLeisOpt	Respondent is asked to choose one of the flights among several alternatives including no choice to resume her leisure travel assuming she has a cover option

Problem	Description
resumeBusOpt	Respondent is asked to choose one of the flights among several alternatives including no choice to resume her business travel assuming she has a cover option



Appendix B

Online Survey Demographic Queries

Uçak Yolculuğu Anket Çalışması

Cinsiyet

Lütfen cinsiyetinizi seçiniz.

Yaş

Lütfen yaş aralığınızı seçiniz.

Aylık Gelir

Lütfen gelir aralığınızı seçiniz.

İleri

Figure B.1: Screenshot of the queries for gender, age and income information

Uçak Yolculuğu Anket Çalışması

Ne kadar sıklıkla iş için uçuyorsunuz?

Uçuşlarınızın yüzde kaçını iş amaçlıdır?

0 % 15 % 100 %

0 10 20 30 40 50 60 70 80 90 100

İleri

Figure B.2: Screenshot of the query for the fraction business travels

Uçak Yolculuğu Anket Çalışması

Yurtiçi **İŞ** seyahatlerinizle ilgili bilgi verir misiniz?

Yurtiçi **İŞ** seyahatleriniz için ödediğiniz minimum ve maksimum bilet fiyatı:

0 TL 250 TL 700 TL 800 TL

Yurtiçi **İŞ** seyahatleriniz için biletinizi en geç ve en erken kaç gün önceden alırsınız?

5 gün 15 gün 60 gün

İleri

Figure B.3: Screenshot of the query for the average minimum and maximum ticket prices for business travels and corresponding purchase times of those tickets

Uçak Yolculuğu Anket Çalışması

Yurtiçi **İŞ DIŞI** seyahatlerinizde ilgili bilgi verir misiniz?

Yurtiçi **İŞ DIŞI** seyahatleriniz için ödediğiniz minimum ve maksimum bilet fiyatı:

0 TL 100 TL 350 TL 800 TL

Yurtiçi **İŞ DIŞI** seyahatleriniz için biletinizi en geç ve en erken kaç gün önceden alırsınız?

0 gün 10 gün 30 gün 60 gün

İleri

Figure B.4: Screenshot of the query for the average minimum and maximum ticket prices for leisure travels and corresponding purchase times of those tickets

Appendix C

Online Survey Query Samples

Uçak Yolculuğu Anket Çalışması

Durum: 5 / 10

İŞ DIŞI SEYAHAT EDİYORSUNUZ

İş DIŞI seyahat ediyorsunuz ve **211** TL ödeyerek aldığınız bir uçak biletiyle beraber aşağıdaki yanmaz bilet opsiyonlarından hangisini almayı tercih ederdiniz?

	Opsiyon A	Opsiyon B	Opsiyon C	Hiçbiri
Opsiyon Fiyatı (TL)	11	21	42	-
Geçerlilik Süresi (Gün)	5	7	10	-

☐ Opsiyon A
 ☐ Opsiyon B
 ☐ Opsiyon C
 ☐ Hiçbiri

Tamam

YANMAZ BİLET OPSİYONU

- Uçağını kaçıran bir yolcunun,
 - kaçırma sebebi ne olursa olsun,
 - havaalanına aynı gün içinde bizzat gelmesi kaydıyla, müsait ilk uçuşta opsiyonun **geçerlilik süresi** içerisinde alacağı yeni biletine **sadece** bilet fiyat farkını ödeyerek sahip olmasını sağlar.
- Opsiyonunuzu geçerlilik süresi içerisinde kullanmazsanız, opsiyon ve yanan biletin ücretleri *iade edilmez*.
- Bu opsiyonu almamışsanız yolculuğunuza devam etmek için, ikinci bilet fiyatının **tamamını** ödemek zorundasınız.

Figure C.1: Screenshot of a query for the optionLeis problem

Uçak Yolculuğu Anket Çalışması

Durum: 1 / 10

İŞ İÇİN SEYAHAT EDİYORSUNUZ

İş İÇİN seyahat ediyorsunuz ve **405** TL ödeyerek aldığınız bir uçak biletiyle beraber aşağıdaki yanmaz bilet opsiyonlarından hangisini almayı tercih ederdiniz?

	Opsiyon A	Opsiyon B	Opsiyon C	Hiçbiri
Opsiyon Fiyatı (TL)	20	61	81	-
Geçerlilik Süresi (Gün)	1	3	10	-

☐ Opsiyon A
 ☐ Opsiyon B
 ☐ Opsiyon C
 ☐ Hiçbiri

Tamam

YANMAZ BİLET OPSİYONU

- Uçağını kaçıran bir yolcunun,
 - kaçırma sebebi ne olursa olsun,
 - havaalanına aynı gün içinde bizzat gelmesi kaydıyla, müsait ilk uçuşta opsiyonun **geçerlilik süresi** içerisinde alacağı yeni biletine **sadece** bilet fiyat farkını ödeyerek sahip olmasını sağlar.
- Opsiyonunuzu geçerlilik süresi içerisinde kullanmazsanız, opsiyon ve yanan biletin ücretleri *iade edilmez*.
- Bu opsiyonu almamışsanız yolculuğunuza devam etmek için, ikinci bilet fiyatının **tamamını** ödemek zorundasınız.

Figure C.2: Screenshot of a query for the optionBus problem

Uçak Yolculuğu Anket Çalışması

YANMAZ BİLET OPSİYONU

- Uçağını kaçıran bir yolcunun,
 - kaçırma sebebi ne olursa olsun,
 - havaalanına aynı gün içinde bizzat gelmesi kaydıyla,
müsait ilk uçuşta opsiyonun **geçerlilik süresi** içerisinde alacağı yeni biletine **sadece** bilet fiyat farkını ödeyerek sahip olmasını sağlar.
- Opsiyonunuzu geçerlilik süresi içerisinde kullanmazsanız, opsiyon ve yanan biletin ücretleri **iade edilmez**.
- Bu opsiyonu almamışsanız yolculuğunuza devam etmek için, ikinci bilet fiyatının **tamamını** ödemek zorundasınız.

Durum: 13 / 20
İŞ DIŞI SEYAHAT EDİYORSUNUZ VE OPSİYONUNUZ YOK

İŞ DIŞI seyahat ederken biletini **233_TL**'ye aldığınız bir uçağı bir aksilik yüzünden yarım saatte kaçırdığınızı düşünün. **Ne yazık ki öncesinde yanmaz bilet opsiyonu almamıştınız.** Aşağıdaki uçuşlardan birisiyle yolculuğunuza devam eder miydiniz?

	Bilet A	Bilet B	Bilet C	Bilet Almam
Bilet Fiyatı (TL)	477	399	360	-
Kalkış Zamanı (Güne Kadar)	1	7	10	-

☐ Bilet A
☐ Bilet B
☐ Bilet C
☐ Bilet Almam

Tamam

Figure C.3: Screenshot of a query for the resumeLeis problem

Uçak Yolculuğu Anket Çalışması

YANMAZ BİLET OPSİYONU

- Uçağını kaçıran bir yolcunun,
 - kaçırma sebebi ne olursa olsun,
 - havaalanına aynı gün içinde bizzat gelmesi kaydıyla,
müsait ilk uçuşta opsiyonun **geçerlilik süresi** içerisinde alacağı yeni biletine **sadece** bilet fiyat farkını ödeyerek sahip olmasını sağlar.
- Opsiyonunuzu geçerlilik süresi içerisinde kullanmazsanız, opsiyon ve yanan biletin ücretleri **iade edilmez**.
- Bu opsiyonu almamışsanız yolculuğunuza devam etmek için, ikinci bilet fiyatının **tamamını** ödemek zorundasınız.

Durum: 3 / 20
İŞ İÇİN SEYAHAT EDİYORSUNUZ VE OPSİYONUNUZ YOK

İŞ İÇİN seyahat ederken biletini **461_TL**'ye aldığınız bir uçağı bir aksilik yüzünden yarım saatte kaçırdığınızı düşünün. **Ne yazık ki öncesinde yanmaz bilet opsiyonu almamıştınız.** Aşağıdaki uçuşlardan birisiyle yolculuğunuza devam eder miydiniz?

	Bilet A	Bilet B	Bilet C	Bilet Almam
Bilet Fiyatı (TL)	640	562	445	-
Kalkış Zamanı (Güne Kadar)	5	7	10	-

☐ Bilet A
☐ Bilet B
☐ Bilet C
☐ Bilet Almam

Tamam

Figure C.4: Screenshot of a query for the resumeBus problem

Uçak Yolculuğu Anket Çalışması

YANMAZ BİLET OPSİYONU

- Uçağını kaçıran bir yolcunun,
 - kaçırma sebebi ne olursa olsun,
 - havaalanına aynı gün içinde bizzat gelmesi kaydıyla,
- müsait ilk uçuşta opsiyonun **geçerlilik süresi** içerisinde alacağı yeni biletle **sadece** bilet fiyat farkını ödeyerek sahip olmasını sağlar.
- Opsiyonunuzu geçerlilik süresi içerisinde kullanmazsanız, opsiyon ve yanan biletin ücretleri **iade edilmez**.
- Bu opsiyonu almadıysanız yolculuğunuza devam etmek için, ikinci bilet fiyatının **tamamını** ödemek zorundasınız.

İŞ DIŞI SEYAHAT EDİYORSUNUZ VE OPSİYONUNUZ VAR

İş **DIŞI** seyahat ederken biletini **162** TL'ye aldığınız bir uçağı bir aksilik yüzünden yarım saatte kaçırdığınızı düşünün. **Şanslısınız ki öncesinde yanmaz bilet opsiyonu almıştınız. Sadece bilet fiyat farkını ödeyerek yolculuğunuza devam edebilirsiniz.** Aşağıdaki uçuşlardan birisiyle yolculuğunuza devam eder miydiniz?

	Bilet A	Bilet B	Bilet C	Bilet Almam
Ödeyeceğiniz Miktar (TL)	315	237	198	-
Kalkış Zamanı (Güne Kadar)	1	7	10	-

☐ Bilet A
☐ Bilet B
☐ Bilet C
☐ Bilet Almam

Tamam

Figure C.5: Screenshot of a query for the resumeLeisOpt problem

Uçak Yolculuğu Anket Çalışması

YANMAZ BİLET OPSİYONU

- Uçağını kaçıran bir yolcunun,
 - kaçırma sebebi ne olursa olsun,
 - havaalanına aynı gün içinde bizzat gelmesi kaydıyla,
- müsait ilk uçuşta opsiyonun **geçerlilik süresi** içerisinde alacağı yeni biletle **sadece** bilet fiyat farkını ödeyerek sahip olmasını sağlar.
- Opsiyonunuzu geçerlilik süresi içerisinde kullanmazsanız, opsiyon ve yanan biletin ücretleri **iade edilmez**.
- Bu opsiyonu almadıysanız yolculuğunuza devam etmek için, ikinci bilet fiyatının **tamamını** ödemek zorundasınız.

İŞ İÇİN SEYAHAT EDİYORSUNUZ VE OPSİYONUNUZ VAR

İş **İÇİN** seyahat ederken biletini **390** TL'ye aldığımız bir uçağı bir aksilik yüzünden yarım saatte kaçırdığınızı düşünün. **Şanslısınız ki öncesinde yanmaz bilet opsiyonu almıştınız. Sadece bilet fiyat farkını ödeyerek yolculuğunuza devam edebilirsiniz.** Aşağıdaki uçuşlardan birisiyle yolculuğunuza devam eder miydiniz?

	Bilet A	Bilet B	Bilet C	Bilet Almam
Ödeyeceğiniz Miktar (TL)	250	172	55	-
Kalkış Zamanı (Güne Kadar)	5	7	10	-

☐ Bilet A
☐ Bilet B
☐ Bilet C
☐ Bilet Almam

Tamam

Figure C.6: Screenshot of a query for the resumeBusOpt problem

Appendix D

The Reference Respondent

Table D.1: The reference respondent demographics for each trip type and their scales

Trip	Demographic	Value	Scale
Leisure	gender2	0	1
Leisure	age2	1	1
Leisure	age3	0	1
Leisure	age4	0	1
Leisure	age5	0	1
Leisure	income2	0	1
Leisure	income3	1	1
Leisure	income4	0	1
Leisure	income5	0	1
Leisure	income6	0	1
Leisure	freq	0.75	1
Leisure	tpMean	250	100
Leisure	tpSpan	250	100
Leisure	ptMean	20	10
Leisure	ptSpan	30	10
Business	gender2	0	1

Trip	Demographic	Value	Scale
Business	age2	1	1
Business	age3	0	1
Business	age4	0	1
Business	age5	0	1
Business	income2	0	1
Business	income3	1	1
Business	income4	0	1
Business	income5	0	1
Business	income6	0	1
Business	freq	0.25	1
Business	tpMean	350	100
Business	tpSpan	350	100
Business	ptMean	12	10
Business	ptSpan	18	10