"Using Competition and Market Information to Optimize a Tour Operator's Pricing Decisions"

Arno Witte

"Using Competition and Market Information to Optimize a Tour Operator's Pricing Decisions"

by

Arno Witte
316985aw

THESIS
for the degree of
MASTER OF SCIENCE

Econometrics and Management Science:
Operations Research and Quantitative Logistics

Faculteit der Economische Wetenschappen
Erasmus Universiteit Rotterdam

ERASMUS UNIVERSITEIT ROTTERDAM

June 2012

Supervisors:
Dr. Adriana Gabor[1]
Dr. Alex Koning[1]
Dr. Kevin Pak[2]

[1]Department of Econometrics, ESE, Erasmus University Rotterdam
[2]Ortec Gouda, Revenue Management Consultant

Abstract

Due to the rising use of internet, the business of e-commerce has grown rapidly in the last years. The easy access to information leads to transparent markets with high competition due to high price sensitive customers. Not only the customers benefit from easy access to information, companies can also benefit from the available market information which is available in many sectors. In this report we propose the Dependent Region Dynamic Pricing (DRDP) model for tour operators. The DRDP model can be used for the dynamic pricing decision of products which make use of both flight and hotel capacities. The DRDP model is an extension of the dynamic pricing model proposed in [50]. The DRDP model incorporates both competition between regional clusters and market information whereas the model from [50] assumes independent regions. The DRDP model makes use of a two-step structure. Therefore, the model is still workable in terms of computation times.

Keywords: •*Revenue Management* •*Online Tour Operator* •*Cross Elasticities* •*Transparent Market* •*Two – Step Model*

Acknowledgements

I would like to express my gratitude to all people who have contributed, directly or indirectly, to the realization of this thesis. I want to thank several people in specific for their contribution. First of all, I am thankful to Adriana Gabor and Alex Koning for their advice and supervision. Adriana is one of the most dedicated and positive minded teachers that I've met in University. I want to thank Kevin Pak for introducing me to the wonderful world of Revenue Management. Your expertise in RM and knowledge about the Sundio project helped me a lot and made me very enthusiastic. Next to this, I am grateful for the fact that you always involved me in new developments and projects. I also want to thank Maarten Soomer for his insights and interest throughout the project. Finally, I want to thank Mark van der Spoel and Ad Peekstok, my direct colleagues at the Sundio project. You guys were willing to help me anytime, without your help I would have lost track in the immense Sundio database.

Arno Witte
Vlaardingen, 22 May 2012

Abbreviation List

AIDS *Almost Ideal Demand System*

BP *Breusch Pagan*

BQP *Binary Quadratic Programming*

DRDP *Dependent Region Dynamic Pricing*

EV *Extreme Value*

GEV *Generalized Extreme Value*

HAC *Heteroskedasticity and Autocorrelation Consistent*

IRDP *Independent Region Dynamic Pricing*

LP *Linear Programming*

MILP *Mixed Integer Linear Programming*

MIP *Mixed Integer Programming*

MNL *Multi Nominal Logit*

MPMR *Multi − Product Multi − Resource*

OLS *Ordinary Least Squares*

RESET *Regression Equation Specification Error Test*

RM *Revenue Management*

VIF *Variance Inflation Factor*

Contents

Part I

Introduction

Chapter 1

Background

Currently, Revenue Management (RM) is widely used in many practical applications in various industries. The most famous applications are those in the airline industry (See for example: [26], [44], [52]), the hotel industry ([20], [25], [27],[35]) and the car rental industry ([16], [39]). According to [9] these are the traditional industries in which RM models are applied. Nevertheless, nowadays many other industries are experiencing the benefits of RM models. Examples can be found in retailing ([4], [38]), media and broadcasting ([8],[33]), tour operating ([50], [42], [43]), cruise ships [28] and casinos [37].

Alongside the growing importance of Revenue Management, the business of e-commerce is also rapidly growing. Moreover, according to [53] the business of e-commerce faced a growth of 18,9% in revenue in 2010. This growth steadily continued in 2011. Typically, in e-commerce there is a constant flow of information which is generated by customers visiting websites. Based on the visitors real-time information, price changes and promotional decisions can be made without human interference. To make proper decisions without human interference, reliable decision models are crucial for a successful business.

Next to the easy data collection, the online setting also adds another dimension to the already complex pricing problems. The important aspect of operating online is that customers can easily compare competing goods through price comparing websites[1]. Next to the travel sector, these price comparing websites are also available in many other sectors. The possibility for customers to compare different competing goods at almost no cost or effort leads to a highly transparent market. In a high transparent market, customers are well informed and highly price sensitive. Examples in the literature are plentiful (see for example: [17], [18] and [41]) for examples of competing traveling regions.

[1] examples of price comparing websites in travel industry are: www.expedia.com, www.orbitz.com, www.travelsupermarket.com

To stay ahead of competitors, it is important to incorporate the latest developments into the decision making process. This can make the difference between success and failure in this fiercely competitive market.

In this report we will focus on improving the RM model proposed in [50]. This RM model is operating in an online setting and has proven to be very successful. In the decreasing market of the economic recession the tour operator was able to transfer 11% more passengers. The realized revenue also increased with 6% and the margin even increased 137%. Despite these impressive results, room for improvement can still be found. We are able to extend this model by incorporating external market information and extended consumer behavior. This results in a robust and more elaborate approach.

1.1 Research Questions

In this study we integrate extended consumer behavior into the RM model proposed in [50]. Beforehand, this results in several research questions. We try to find the answers to these questions. Some of these questions raise new sub-questions. We will also define some additional questions which are relevant for this study. Let us first propose the following main research question (RQ).

1.1.1 Main Research Question

RQ: *How can we integrate extended consumer behavior and market information into a tour operator's RM model?*

1.1.2 Research Sub-Questions

In this study we use a data driven approach, this means we focus on the observed consumer behavior. The first thing we want to know is how we can integrate the market information into the tour operators pricing decision. We are also interested in how we can measure the consumer behavior from historical booking data. After we measured the observed behavior, we want to know how this can be integrated into a RM model. Therefore, this question can be split into three sub-questions (SQ.1 - SQ.3).

SQ.1: *How can we integrate market information into the tour operators pricing decisions?*

SQ.2: *How can we measure the consumer behavior from the historical booking data?*

SQ.3: *How can we integrate the measured consumer behavior into a RM model?*

In Chapter 4 we focus on measuring the consumer behavior from the historical booking data. We focus on the behavior of customers after price changes compared to the market price. In Chapter 5 we focus on integrating this consumer behavior into a RM model.

1.1.3 Other Relevant Questions

Next to the research questions and sub-questions we are also interested in other aspects of the proposed RM model. Therefore, we also try to answer general questions about the validity and practicality of the model. It is important to have answered these questions before the model can be used in practice.

- *What is the validity of the proposed model?*
- *How can we measure this validity?*
- *How does the extended model perform compared to the original model?*
- *Is the extended model still workable in terms of computation times?*
- *Are the additional computation times acceptable for practical application?*
- *Are the expected benefits interesting for practical usage?*

1.2 Outline

In this section, a brief summary of the following chapters is provided. In this first chapter, we provided some background on RM models and the important online setting. We also stated the research questions, these questions will be answered in the remainder of this study.

In Chapter 2, some insights in the actual problem are provided. We provide an overview of the complete decision making process for the online setting of the pricing problem. This overview enables us to place the RM model in the bigger picture of the tour operator's decision making process. One of the most important things is the composition of the travel products, the products the tour operator offers. We see that these products are competing for space among multiple flights and hotels, this results in a considerable combinatorial problem. [50] tackled this problem using an Independent Region Dynamic Pricing (IRDP) model. We also focus on the possible limitations of the IRDP approach.

In Chapter 3, the relevant literature is studied extensively. This literature study helps us placing our research into perspective.

In Chapter 4, the observed consumer behavior is measured and quantified. The consumer behavior is captured in a multiple regression model. This multiple regression model includes cross regional effects and the external market information. The goal of the regression model is to estimate the expected market share for a a certain region, for a given pricing decision.

In Chapter 5, the quantified consumer behavior is integrated into a Dependent Region Dynamic Pricing (DRDP) model. The DRDP model contains two stages: in the first stage the regional cross effects are used to identify the optimal region price, this price is taken into account while pricing the individual products in the second stage.

In Chapter 6 we provide an analytical comparison between the IRDP model and the DRDP model. First, we focus on the problem size of the different models. Next, we focus on the comparison of the objectives of the different models.

In Chapter 7 we provide a case study on Sunweb, a large online tour operator based in the Netherlands. The cross regional effects are measured using actual booking data. Hereafter, the cross regional effects are used to solve the DRDP model for multiple problem instances. For all instances, we use simulation to compare the results of the DRDP pricing decision to the IRDP pricing decision.

Finally, In Chapter 8 we draw the main conclusions and we provide some managerial advice.

Chapter 2

Problem Description

In this chapter we provide a detailed description of the research problem.

In section 2.1, the setting of the problem is described. The RM model proposed in [50] can be generally applied to a tour operator facing the described problem setting. Important is the structure of the travel products. These travel products contain the unique combination of several properties. Subsequently, an overview of the decision making process for the online setting of the problem is provided. This overview enables us to place the RM model in the complete picture of the decision making process.

In section 2.2, we focus on the step of the RM model in which the prices are determined. The pricing model is based on the assumption that the regions are independent, therefore we refer to this model as the IRDP (Independent Region Dynamic Pricing) model. The assumption of independent regions considerably simplifies the problem.

In section 2.3, possible limitations of the IRDP model are discussed. We describe the way the consumer behavior is modeled in the IRDP model. We also describe the consumer behavior in reality. The goal of this research is to close the gap between the consumer behavior in the IRDP model and the consumer behavior in reality. We try to find a RM model that describes reality as close as possible.

2.1 Required Problem Setting

The RM model proposed in [50] is tailor made for a specific tour operator. Nevertheless, this model can be generally applied to any tour operator that faces a similar problem setting. In this section we describe the problem setting that is required to apply the proposed RM model successfully. An important requirement is that the composition of the travel products must have a certain structure. This structure is described in 2.1.1. Subsequently, the specific requirements for the decision making process are described in 2.1.2.

2.1.1 Composition of the Travel Products

The first requirement of the problem setting has to do with the composition of the travel products. Each product consists of a unique combination of the following properties:

- Brand
- Destination
- Accommodation
- Room Type
- Start Date
- Outbound Flight
- End Date
- Inbound Flight

Sometimes, tour operators are operating using different brand names. In that case, the brand is used to differentiate products. All products contain a certain destination. Next to this destination, a product consists of the stay, inbound and outbound flights for given dates. This means that the products with the same destination are competing for space among multiple flights and accommodations. For example, products competing for the same outbound flight might be booked to a different accommodation with a different duration. This leads to a large number of products using the same outbound flight. The same holds for inbound flights and for the accommodation capacity. A product consists of a unique combination of its attributes, this can lead to a very large number of unique combinations, this is illustrated in example 2.1.

Example 2.1. *Consider two travel products departing the same day. The duration of the trips is 8 and 11 days respectively. Both products can make use of the same inbound flight, so the inbound flight overlaps. Next to this, both products can make use of the same accommodation, in that case the stays overlap 8 accommodation days. In total, the 8 day trip influences ten different capacities. The 11 day trip influences different 13 capacities. Note that when the number of products grows, the number of overlaps also grows.*

Example 2.1 clearly shows that there are a high number of products influencing the same capacities. This means that if a product is booked, the capacities of many other products are also influenced. This example shows the overlap between two travel products, in real life there might be millions of travel products. Because of the overlap, only a small percentage is actually booked. The overlapping capacities of the travel products generate a considerable combinatorial problem. The challenge is to price these products subject to these capacity constraints to generate maximum revenue. In section 2.2 we discuss how [50] tackled this complex problem.

2.1.2 The Decision Making Process

Other requirements for the problem setting are related to the decision making process. First of all, let us give an overview of the general decision making process a tour operator faces. According to [47] the decision making process of a tour operator consists of three general stages.

1. Determine start capacities:
The first stage consists of the capacity planning, in this stage the starting accommodation and flight capacities have to be determined. These capacities are usually outsourced to third-party suppliers. Also, in the travel industry the initial capacities are often assumed to be fixed. This means that the flight and hotel capacities are reserved beforehand. Due to this assumption we can treat the available capacities as given; this considerably simplifies the pricing decisions.

2. Determine base prices:
In the second stage, based on the determined capacities, a base price is determined for each of the products. These base prices are adjusted in the RM stage to respond to recent developments.

3. RM model/ dynamic pricing:
The third stage is considered as the RM stage, in this stage the revenue is maximized as a reaction to the operational information obtained from the tour operators websites. In this study we mainly focus on this third stage considering the input from the other stages as given. In figure 2.1 we see the diagram of the tour operator's decision making process. The shaded areas are the stages which are under control of the tour operator, the three general stages from [47] can clearly be distinguished.

In figure 2.1 we extended the decision making process for an online tour operator. This figure is quite similar to the figure presented in [43], they also focus on the decision support for online travel retailing. We see the population of potential customers visiting the websites of the tour operator. There are n brands, all brands have their own website on which they present the travel products. The different brands might be focusing on different market segments, for example youth, elderly people or families. The potential customers might also be visiting the websites of competing tour operators. From the websites the operational information is gathered. The operational information is used to make long and short term forecasts for the number of bookings. From the long term forecasts, long term deals can be closed with third party suppliers. The long term forecasts are also used

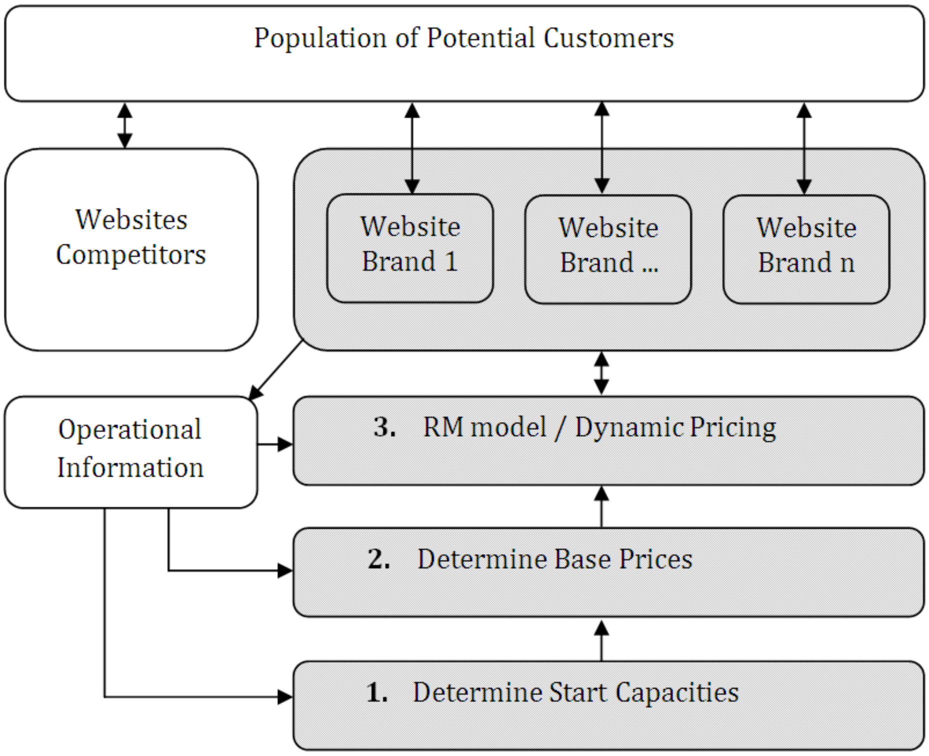

Figure 2.1: Overview of the decision process for an online tour operator

to determine the base price for each of the products. From the short term forecasts the capacity planning is made and the prices are optimized within the RM model. In reality it is sometimes possible to buy additional capacities at third party suppliers. In this case the capacities entering the RM model have to be adjusted. Finally, prices determined in the pricing model are displayed on the websites.

2.2 Formulation of the IRDP Model

The Independent Region Dynamic Pricing (IRDP) model is the step in the RM model in which the prices for the different products are determined. The IRDP model is an example of a Multi-Product Multi-Resource (MPMR) pricing model. In such a model the prices of multiple products which make use of multiple resources are optimized. These resources can for example be airline seats or hotel rooms. In this section we will give a detailed description of the IRDP model. We first define the used sets, parameters and variables. Hereafter, the actual IRDP model is formulated.

Sets:

$i \in I$:	Set of products
$j \in J_i$:	Set of candidate prices defined for product i
$t \in T$:	Set of time periods
$k \in K$:	Set of resources
$S_k \subseteq I$:	Set of products making use of resource k

Let I be the set containing all products i. For each product we define set J_i containing candidate prices near the current price. The booking period T for each of the products is partitioned into t smaller time intervals. The time intervals can be chosen such that the arrival rate of purchasing customers is assumed to be constant in each time interval. Let K be the set of resources that can be used by the products. Note that one product can make use of multiple resources, but also that a resource can be used by multiple products. For each of the resources k we define the set $S_k \subseteq I$ of products that make use of this resource.

Parameters:

p_{ij}:	Candidate price j for product i
c_i:	Cost price of product i
q_{ijt}:	Expected demand for product i if candidate price j is set in time period t
Cap_k:	Available capacity for resource k

Let p_{ij} be the price for price candidate j of product i and let c_i be the cost price of product i. q_{ijt} is the expected number of customers purchasing product i when price j is set in period t. Price elasticities are incorporated for determining q_{ijt} at different price levels. These price elasticities are measured from the historical operational information for the different products. We also define Cap_k as the available capacity of resource k at the beginning of the optimization period.

Variables:

x_{ijt}:	Fraction of time price j is chosen for product i in interval t

The goal of the IRDP model is to maximize the expected revenue over the defined time period T. In the pricing decision we have to decide which price to choose for each product in each time interval. To do this, we introduce decision variables $x_{ijt} \in [0, 1]$. x_{ijt} is defined as the fraction of time interval t price j is chosen for product i. The price fractions form a price path for every time interval. The resulting price path is used such that the price path is decreasing (i.e. the highest price with a non-zero price fraction is chosen first and the lowest price with a non-zero price fraction is chosen last). The optimal price fractions are determined using the Linear Programming (LP) model in (2.1)-(2.4).

Maximize:

$$\pi_{IRDP} = \sum_{i \in I} \sum_{j \in J_i} \sum_{t \in T} (p_{ij} - c_i) q_{ijt} x_{ijt} \tag{2.1}$$

Subject to:

$$\sum_{i \in S_k} \sum_{j \in J_i} \sum_{t \in T} q_{ijt} x_{ijt} \leq Cap_k, \quad \forall k \in K \tag{2.2}$$

$$\sum_{j \in J_i} x_{ijt} \leq 1, \qquad \forall i \in I, t \in T \tag{2.3}$$

$$0 \leq x_{ijt}, \qquad \forall i \in I, j \in J, t \in T \tag{2.4}$$

The objective of the IRDP model sums over all periods, products and possible candidate prices. For each period, the revenue for each product $(p_{ij} - c_i)$ is multiplied by the expected demand for this price path $(q_{ijt} x_{ijt})$. The objective of this model is to choose the price path for each product that maximizes the total expected revenue over the given time horizon T.

Using constraint (2.2), we make sure that for each of the resources $k \in K$ the demand does not exceed the available capacity of this resource. Constraints (2.3) and (2.4) are constraints on the price path. These make sure that the price path meets several conditions. Constraint (2.3) ensures that for each time period the price path fractions are smaller or equal to one. Additionally, in constraint (2.4) we make sure that all fractions are non-negative.

For practical use, the IRDP model runs in a quasi-static environment. This means that price changes are made every night based on the information received during the day. The demand forecasts, capacities and elasticities can also be updated in an overnight process. The data on which these computations are based are stored in a rolling horizon containing the historical data.

2.2.1 Decomposition of the IRDP model

In this subsection we describe the main practical advantage of the IRDP model. In the IRDP model, products to different regions do not use overlapping capacities. Therefore, the problem can be split into smaller sub-problems per region. This can be done using the decomposition described in this subsection.

First, note that the LP model in (2.1)-(2.4) can be written in standard form (2.5).

$$\begin{aligned} \max \quad & \pi = \mathbf{cx} \\ \text{s.t.} \quad & \mathbf{Ax} \leq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}, \end{aligned} \tag{2.5}$$

with $\mathbf{A}$ the mn-matrix of constraint parameters.

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix},$$

$\mathbf{x}$ is the column n-vector of decision variables, $\mathbf{c}$ is the column n-vector of cost coefficients and $\mathbf{b}$ is an m-vector with non-negative right hand side parameters. Also note that $\mathbf{0}$ is the column n-vector of zeros.

First let us introduce set R containing all travel regions. Usually travel products to different travel regions do not make use of the same hotel or flight capacities. When we assume that the pricing decision for one region does not influence the demand for other regions we can treat the regions independently. This means that the set of products and the sets of resources can be partitioned into subsets. In other words, we can split the set of products I into subsets of products to each region $I = \{I_1, I_2, \dots, I_r\}$.

By partitioning the set of products into the subsets of products per region, the LP model can be written for each region separately. Let $\mathbf{c_r}$, $\mathbf{x_r}$, $\mathbf{A_r}$ and $\mathbf{b_r}$ be the vectors corresponding to region r. The problem can be written as the following r sub-problems:

$$\begin{aligned} \max \quad & \pi_{\mathbf{r}} = \mathbf{c_r x_r} \\ \text{s.t.} \quad & \mathbf{A_r x_r} \leq \mathbf{b_r} \qquad , \forall \mathbf{r} \in \mathbf{R} \\ & \mathbf{x_r} \geq \mathbf{0} \end{aligned} \tag{2.6}$$

Definition 2.1. *When the sub problems in (2.6) are independent and π_r is the optimal solution for sub problem r. Then, the optimal solution of the overall problem in (2.5) (π) can be expressed as the sum of the optimal solutions of the sub problems. $\pi = \sum_{r \in R} \pi_r$.*

Usually the complexity of a LP model grows more than linearly in the number of decision variables. Therefore, decomposing the LP problem into r independent sub problems considerably reduces the complexity of the problem.

2.3 Limitations of the IRDP Model

In the introduction we described the satisfactory results realized after implementing the IRDP model from [50] at an online tour operator. Nevertheless, room for improvement can still be found. In this section we will focus on possible limitations of the IRDP model.

First, we look at the way consumer behavior is modeled in the IRDP model. Next, we look at consumer behavior in reality. In the remainder of the report we try to model this consumer behavior as closely as possible. The final goal is to improve the RM model proposed in [50], so that it more accurately describes the actual consumer behavior.

2.3.1 Consumer Behavior in IRDP Model

In the IRDP model in (2.1)-(2.4), price elasticities are used to predict the demand for a given product for different price alternatives. Thus, in the current approach the only trigger for additional customers to buy a certain product is the price change of the given product. In this situation, we assume that customers are not interested in substitutable products. According to various literature sources ([14], [17]-[19], [47]) there are strong correlations between different travel products. [19] states that the Internet has brought consumers increased access to information to make purchase decisions. As markets come closer to perfect information, one of the expected outcomes is an increase in competition. Therefore, we expect that the negative consequences of the simplifying assumption of independent regions increase in the current market.

Another limitation of the current IRDP model is that external market shocks (other than seasonality) are not accounted for in the elasticity estimation, see example 2.2.

Example 2.2. *The Arabic Rebellions starting in 2010 led to a decrease in tourism in Egypt and Tunisia. In the current model this decrease might be wrongly attributed to price changes. Thus, when we use historical data of two years to estimate the price elasticities for these countries, the estimates will be biased for the coming two years.*

Despite the fact that this is a rather extreme example, it shows that in general each external market shock (or shift) possibly causes biased elasticity estimates. Numerous examples of such shocks can be found in reality. The most important thing to note is that as long as these shocks are part of the historical dataset, the elasticities remain biased.

2.3.2 Consumer Behavior in Reality

In reality, customers will choose a product based on numerous interacting factors. We aggregate these factors in the following categories: price changes of the product, price changes of substitutable products, market shocks and other external factors influencing demand. Next, we explain why each of these groups might influence the demand for a given product. The challenge of this research is to account for these factors in the pricing decisions. In other words, the challenge is to formulate a pricing model in which these factors are incorporated.

Price Changes of the Product

Obviously, the price of a product influences the demand for the product. A higher price will lead to lower demand and a low price will lead to higher demand. This leads to the well-known supply and demand curves from which the price elasticity of a product can be estimated. The price elasticity of a product is already incorporated in the IRDP model.

Price Changes of Substitutable Products

In addition to the price changes of the product itself, prices of substitutable products can also influence the demand for a product. When the prices of substitutable products fall, more people tend to choose the substitutes over the given product. On the other hand, when the prices of the substitutes rise more people are tend to choose the given product.

The substitutable products category is divided into two sub-categories: substitutable products within the tour operator's products and substitutable products offered somewhere else in the market.

Within the first sub-category the tour operator fully controls these prices. It might be possible for the tour operator to influence the customers to choose certain products with large margins or available capacities.

The second sub-category consists of competing with products from other tour operators in the market. Since tour operators operate in a highly transparent market ([17] and [18]), the customers are well informed and highly price sensitive. Customers who are willing to buy a certain travel product will look in the market for substitutable products and choose a product with a favorable price. This is illustrated in example 2.3:

Example 2.3. *A customer is interested in a holiday trip for a certain date. He finds an eight day trip to Crete, what he likes about Crete is the sun, the sea and the beach. After he found this trip, he searches the web for comparable trips to the Aegean Coast. When he finds out that trips to Rhodes are less expensive he is triggered to book a 8 day trip to Rhodes instead.*

This illustrates the price sensitivity of the customers and also the competitiveness of products with comparable properties. Therefore, it is also important to look at the prices of products compared to the prices of substitutes in the market.

Shocks in Market Demand

Another direct effect in demand shocks come from shocks in the complete market. All shocks that influence the total market demand are part of this category. Example 2.2 also illustrated a shock in market demand. Market shocks can for example be caused by: seasonality, economic recession, terrorist attacks, war or natural disaster. Numerous other examples can be found. Most of the time, it is hard to predict shocks in market demand. However, we can define robust models in which the market shocks are less influential.

Other External Factors

The shocks that are not due to price changes or market shocks are categorized as other external shocks. These shocks can for example be caused marketing activity of the tour operator or by negative publicity for the tour operator or one of its brands. In general these other external factors cause shocks that influence the tour operator in a different way than the total market.

Only the price changes of the products are incorporated in the consumer behavior in the current RM model. In this research we focus on incorporating price changes of substitutable products into the elasticity estimation. In the next chapter we will discuss relevant literature. This might provide useful ideas and methods for measuring the consumer behavior and for extending the IRDP model. This also enables us to place our research within the existing literature.

Chapter 3

Related Literature

Before we can focus on the RM model we first have to focus on measuring the cross price effects. Current literature provides numerous estimation techniques that can be used to estimate the price and demand relation between multiple products. For example, in retailing a lot of research is done on cross price effects ([10], [24], [45], [15] and [29]). The latter two are the most interesting (and recent), they are discussed below. Most of this research is based on store level scanner data[1] or survey data. The scanner data consists of the aggregate purchases for given periods (e.g. the weekly purchases for a given product) and is therefore very similar to the historical booking data of a tour operator.

In [15], scanner data from multiple stores is used to estimate the (cross) price effects between private labels and national brands of groceries. They propose a log-log regression model to estimate the demand-price elasticities. Instrumental variables are used to account for price endogeneity. Like in [10], [24] and [45] the wholesale price of a product is used as an instrument. Although the use of instrumental variables seemed to work well, the wholesale price is not directly available for tour operators. This is because the capacities are bought at the beginning of the season for a fixed price. In addition, travel products make use of multiple capacities which makes it more difficult to determine the exact wholesale price of a product. However, in our research, before estimating the cross price effects we perform a data transformation which reduces the endogeneity problem.

[29] investigate the cross category demand effects of promotions. They are also interested in the cross price effects of substitutable products. To avoid a high number of cross product effects, they divide the products into different independent clusters. These clusters are again divided into different categories between which category effects are measured. This clustering mechanism can also be used by a tour operator. The travel products can

[1]This is Purchase information (for example: price, brand, product size, amount purchased) gathered at the time of purchase by an electronic device.

also be divided into independent clusters (e.g. sun products, winter products). Within these clusters we also create sub categories, for example the countries, regions or destinations of the different products. In our case the travel products are very specific, most of the products are never booked at all. This also makes it necessary to aggregate the booking data on a higher level. In the case of a tour operator, the geographical properties of products offer a natural way to cluster the products. In our research we will cluster the products on a higher regional level.

Also plenty literature is available on measuring price and cross price elasticities between different holiday destinations ([12], [18], [19], [34]). A frequently used model is the Almost Ideal Demand System (AIDS) model proposed in [11]. In [12] this model is developed for the tourism context. In the AIDS model, the budget share for travel destinations is estimated from the average expenditures to these destinations. An interesting application of the AIDS model is found in [34], they propose a hedonic pricing model to explain the effects of product characteristics on the products prices. The problem with this approach is that expenditure data of the different regions is not always directly available. Besides this, updating the estimates from this model needs new expenditure data. Therefore, for a tour operator's pricing decision these models are not useful. Nevertheless, these studies provide insight in the expenditures of customers and price competitiveness between different destinations.

In most cases, elasticities are estimated using a regression model. A comprehensive and detailed overview of regression techniques, assumptions and validity tests is given in [23]. We will also use these techniques to model the consumer behavior.

There is also plenty literature available on RM models. We refer to [5], [47] and the references therein for an overview. Particularly, we refer to [14] for an overview of RM models under inventory considerations. However, not much of this literature proposes RM models specifically for tour operators. Besides this, pricing models from the airline and hotel industries are mostly not directly applicable for a tour operator. The combinatorial effects resulting from products competing for limited hotel and airline capacities result in a completely different problem structure. Therefore, as stated in [50], RM models from the airline and hotel industry do not fit the tour operator business properly.

Like in this study, some recent studies incorporate consumer behavior in RM models ([3],[6],[13],[38]). They all use consumer choice models to model the customers preferences. A frequently used consumer choice model is the Multinomial Logit (MNL) model, this is used in [3] and [13]. But also other consumer choice models are studied, in [6] the customers buying decisions are modeled using a knapsack problem. In [38], products that are likely to be bought simultaneously are offered as a package of products. Most of these studies focus on buying decisions of customers that possibly buy multiple items. In the travel industry customers will not buy multiple travel products at once, this has the consequence that these models not directly applicable in the travel industry. However, in the travel industry people can be influenced to buy a certain travel package with a high margin or large free capacities.

For tour operators, several researchers mention the necessity of taking into account the cross price elasticities (for example, [14] and [47]). Despite this, we do not know of any research that actually integrates cross price decisions into a tour operator's dynamic pricing model.

Part II

Methods

Chapter 4

Statistical Methods

In this chapter we discuss how the more advanced consumer behavior from section 2.3.2 is measured and quantified.

First, in section 4.1 the market share functions and the relative price functions are defined. We use the market shares and relative prices to model the characteristics of a transparent market in which customers can easily compare prices. We also propose a multiple regression model, used to measure the cross elasticities. In the multiple regression model, we use the relative prices of a region compared to the market and other regions to explain the market share of the region.

Next in section 4.2, we propose a simulation technique that can be used to estimate the price elasticities of demand from the estimated cross elasticities. We simulate a dataset and from this we estimate the price elasticities of demand using a simple regression model.

A (multiple) regression model is based on several regression assumptions. In section 4.3 we discuss the process of validating and testing these assumptions.

4.1 Cross Elasticities

In section 2.3.2 we stated that the demand shocks are caused by price changes, price changes of substitutes, market shocks and external shocks. We will present and approach in which price changes and price changes of substitutes are incorporated into the decision making process. Our approach is also more robust in case of market shocks. The (cross) price effects are modeled using a multiple regression model. In this section we propose a method to estimate the cross price effects from the historical booking (and market booking) data. The cross product effects are measured on the region level. The products are clustered on regional level, as described in section 2.2.1. In the clustering decision we took into account the fact there has to be enough data available on cluster level to measure effects. The regional level is also chosen because of the decomposition property which is used again later on, in the optimization. After we defined the regional clusters, we can measure the cross regional effects between these clusters. First, let us introduce the used definitions for market share and relative prices. We use capital letters to denote the properties of the regional clusters.

4.1.1 Market Share and Relative Prices

Let Q_{rt} be the demand for the products to region r in time period t. First, let us define the market in definition 4.1:

Definition 4.1. *The market is defined as: all parties that directly compete with the tour operator by offering similar products in the same market.*

We chose to define the market as only direct competitors. It is important that the market consists of actual competitors, else the desired cross product effects will be disturbed.

Let M_t be the total market purchases in time period t. Let S_{rt} be the tour operators market share for region r in time period t. The market share is defined as the fraction of purchases to region r at tour operator relative to the total market purchases. This relationship is defined in (4.1):

$$S_{rt} := \frac{Q_{rt}}{M_t}. \tag{4.1}$$

For each region, we also define the price of travel products relative to the average market price. Let P_{rt} be the price for the products to region r in time period t. Next, let N_{rt} be the market price for the products to region r in time period t. Now, the relative price G_{rt} is defined as the average price of the tour operators products to region r compared to the average market price for region r. This relationship is defined in (4.2):

$$G_{rt} := \frac{P_{rt}}{N_{rt}}. \tag{4.2}$$

We assume that the potential customers do not only compare other tour operators, but also to other comparable regions. Therefore, we also define the relative price of region r compared to other region o. The relative between region r and other region o at time t is defined in 4.3:

$$H_{rot} := \frac{P_{rt}}{P_{ot}}, \quad \forall o \neq r. \tag{4.3}$$

In this study we focus on the market share series of a region rather than on the actual demand series of the region. Dividing by the market demand series filters the series from all patterns which are also present in the total market, this property has a lot of advantages. For example, the series is filtered from seasonal patterns. Next to this, the series is also filtered from market shocks. This makes the elasticities more robust, see example 4.1.

Example 4.1. *Let us return to example 2.2 of the Arabic Rebellions (section 2.3.1). The Arabic Rebellions caused a decrease in tourism to Egypt and Tunisia in the total market. The market demand decreased enormously. However, the distribution of the market over the tour operators is not expected to change. So, although there is a shock in demand, there is no shock in market share. Therefore, elasticities based on market share are more stable in this case than elasticities based on the actual demand.*

In example 4.1 we illustrated that the market share is expected to be much more stable over time than the demand. By using the market shares, we indirectly correct for external shocks in the market. This way we correct for seasonality, economic recession, terrorist attacks, war, natural disaster and numerous other factors present in the entire market. Therefore, the elasticities based on the market share will also be more stable than the elasticities based on the actual demand.

We also focus on the relative prices, instead of the actual prices. Higher absolute prices are not necessarily a good benchmark for lower demand, see example 4.2.

Example 4.2. *When the 'season' starts in January, the market prices are relatively high. The market demand is also relatively high, because a lot of people book their holidays in January. The opposite is true for the end of the season, there are low 'last minute' prices and low market demand. In a transparent market, the relative prices of different tour operators compared to each other are expected to move the market to the cheapest tour operator. The relative prices of different traveling regions are expected to move the market to the cheapest regions.*

This example illustrates why it is might be a good idea to look at the relative prices instead of the actual prices. The customers are not necessarily triggered by low prices, they are triggered by low relative prices. The market tends to shift to the tour operators and regions with the lowest relative prices at a given time. By using the relative prices we incorporate the consumer behavior present in a transparent market. In a transparent market, customers can easily compare the products (and prices) to other products in the market. This feature can be modeled by using relative prices.

4.1.2 Multiple Regression Model

In this subsection we propose an Ordinary Least Squares (OLS) multiple regression model to estimate the cross elasticities between different regional clusters. We use the definitions of the market share (4.1), the relative price compared to the market (4.2) and the relative price compared to other regions (4.3).

We try to explain the market share S_{rt} of region r at time t with the relative price at this time compared to both the market (G_{rt}) and other traveling regions (H_{rot}). However, we do not expect that all regions relative prices will significantly influence the market share of the given region. Only the relative prices of highly substitutable regions are expected to influence each other's market shares. Therefore, we use a backward elimination strategy[1] to determine which regions relative prices influence the market share of the given region significantly. We start using the regression in (4.4):

Regression Model:

$$S_{rt} = \alpha_r + \beta_{rr} G_{rt} + \sum_{\substack{o \in R \\ o \neq r}} \beta_{ro} H_{rot} + \varepsilon_{rt}, \;\; \forall t \in T. \tag{4.4}$$

The parameters from the model in (4.4) can be computed using general OLS formulas. Amongst others, these formulas can be found in [23]. The estimates ($\widehat{\boldsymbol{\beta}}$) can be interpreted as the sensitivity between the relative prices and the market share, this relation is expected to be negative. When the tour operator increases the relative price for region r the market share of r is expected to decline.

In the backward elimination, we eventually defined the cross regional relationship between all regions that have a significant regression coefficient. If the coefficient between relative price and market share is not significant, no relationship between the regions is found. Estimating the coefficients we must keep in mind that our goal is not the overall fit of the model, our aim is to calculate the correct elasticities.

Eventually, the regression equation can be used to express the expected demand as a function of the given pricing decision. This expression can be found in equation (5.6) in chapter 5.

4.2 Elasticities

In this section we describe how we can use the cross elasticities from the previous section to calculate the individual price elasticities of demand. In the IRDP model, the individual elasticities of demand are required to determine the expected demand after a price change for an individual product. Often for a tour operator it is very hard to estimate the price elasticities of demand from the booking data, see example 4.2. Using the regression equation (4.4), we can calculate the expected market shares per region from given prices per region. When we multiply this with a given market demand we obtain the expected demand per region. This relates the price and demands between the regions. When we simulate prices from estimated price

[1] A commonly used strategy in regression analysis

disrtibutions and market demand from the estimated market demand distribution, we can calculate the demand per region. This way, we can create a dataset from which we can estimate the price elasticities of demand.

4.2.1 Price and Market Demand Distribution Fits

We use the actual data to fit distributions of the market demand and the prices. We use a dataset of market demands, and a dataset of prices (the tour operator's prices and market prices). To see which distribution fits the data properly we use the Kolmogorov-Smirnov (KS) test. The KS test is based on the maximum difference between an empirical and a hypothetical cumulative distribution. It is widely used to determine the "goodness of fit" of a theoretical distribution compared to the distribution of the sample values (for more information about the KS test we refer to [36]). We try multiple different distributions [2], for each distribution we estimated the best fitting parameters using maximum likelihood. After this, the p-value of the KS test is used as a measure to determine which distribution fits the best to the sample data. First we describe two distributions that seem to fit well to market share and price data.

Extreme Value Distribution

The EV distribution used as a model for extreme values, it can also be used as a model for other types of continuous data. For example, EV distributions are closely related to the Weibull distribution. If the data has a Weibull distribution, then log of the data follows an EV distribution. The pdf of the EV distribution is given in (4.5).

$$f(x) = \frac{1}{\sigma}exp(\frac{x-\mu}{\sigma})exp(-exp(\frac{x-\mu}{\sigma})) \tag{4.5}$$

Generalized Extreme Value Distribution

The Generalized Extreme Value (GEV) distribution covers three types of distributions, these are called type I, type II and type III. Type I covers the type of distributions with exponentially decreasing tails, such as the normal distribution. Type II covers the distributions with polynomial decreasing tails, such as the student t distribution. Type III covers the distributions which tails are finite, such as the beta distribution. By using maximum likelihood to estimate the parameters of the GEV distribution we let the data decide which type of distribution is appropriate. The pdf of the GEV

[2] We tried the following distributions: Birnbaum Saunders, Exponential, Extreme Value, Gamma, Generalized Extreme Value, Inverse Gaussian, Logistic, Log-Normal, Nakagami, Normal, Rayleigh, Rician, Student-t and Weibull

distribution is given in equation (4.6).

$$f(x) = \begin{cases} \frac{1}{\sigma}(1 + \xi(\frac{x-\mu}{\sigma}))^{(-1/\xi)-1} exp(-(1 + \xi(\frac{x-\mu}{\sigma}))^{1/\xi}\}, & \xi \neq 0 \\ \\ \frac{1}{\sigma} exp(-(\frac{x-\mu}{\sigma}) - exp(\frac{\mu-x}{\sigma})), & \xi = 0 \end{cases} \tag{4.6}$$

where μ, σ and ξ are the location, scale and shape parameters respectively. Furthermore, it must hold that $\sigma \leq 0$ and $\mu, \xi \in \mathbb{R}$. According to [32], the determination of the paramater ξ is the central problem of extreme value analysis. When $\xi > 0$, the underlying distribution belongs to the Fréchet maximum domain of attraction and is regularly varying (power-like tail). When $\xi = 0$, it belongs to the Gumbel Maximum Domain of Attraction and is rapidly varying (exponential tail), while if $\xi < 0$ it belongs to the Weibull Maximum Domain of Attraction and has a finite right endpoint.

4.2.2 Creating the Dataset

For each region we draw a random number from the fitted price and market demand distributions. We then have the price, market price and expected market demand for each region. Using the cross elasticity fits from formula (4.4) we can compute the expected demand for each region. This procedure is repeated multiple times to create a dataset of sufficient points. The price elasticities of demand can be estimated from the prices with corresponding demands for each region, using a simple OLS regression model.

4.3 Statistical Tests

Using a (multiple) regression model, we indirectly make several statistical assumptions about the predictor variables, the response variables and their relationship. Among others, these assumptions are described in detail [23] (section 3.1.4).

Not meeting these assumptions, provides several pitfalls that possibly undermine the validity of the estimates. In case all assumptions hold, the analytical accuracy methods used in the regression are valid. Moreover, these assumptions imply that the parameter estimates will be unbiased, consistent, and efficient in the class of linear unbiased estimators. Therefore, it is important to take notice of these assumptions. Finally, in the multiple regression we make the assumption that there is no multicollinearity. Multicollinearity would cause difficulty in determining the individual effects of the independent variables. We are mainly interested in these individual effects, therefore this assumption is also important.

4.3.1 The Regressors

1. Fixed Regressors

This means that the explanatory variables are assumed to be non-random, describing the situation of controlled experiments. In our case, we consider a real data example. Therefore, we assume that there are no external factors that influence the chosen variables. Influential data points that cannot be explained by the model are detected in the outlier detection mechanism. These data points are omitted.

4.3.2 The Disturbances

2. Random Disturbances, zero mean

The disturbances from the regression model are random variables with zero mean. We can use an ANOVA to test the zero mean assumption, the disturbance distribution is tested in assumption 7. When the mean of the disturbances is not zero, the model structurally overestimates or underestimates the dependent variable. This means that the estimates are biased.

3. Homoskedasticity

We also assume that the error terms of the model have a constant variance (i.e. the error terms are homoskedastic). When this is not the case, the OLS estimates can be inefficient. We use the Breusch-Pagan test to test for homoskedasticity in the error terms. The Breusch-Pagan test checks whether the estimated variances of the residuals from a regression are dependent on the values of the independent variables. A way to account for heteroskedasticity is the use of Weighted Least Squares (WLS). In WLS, each observation is given a weight in determining the estimates $\widehat{\beta}$. The idea is that observations with a smaller variance are more certain and hereby get a larger weight.

4. No Correlation

Another assumption for applying OLS is that there is no correlation present in the error terms (i.e. ε_i and ε_j are not correlated for all $i \neq j$). We use the Breusch-Godfrey test to test whether previous p error terms influence the current error term significantly. This is done using an autoregressive model of order p (commonly referred to as $AR(p)$ model) as an auxiliary regression. From this auxiliary regression we can test whether serial correlation between the error terms is present. A common solution to correlated error terms is the use of Newey-West standard errors. The standard errors of the OLS estimates using Newey-West standard errors are considered HAC (Heteroskedasticity and Autocorrelation Consistent). Therefore, by using newey-west standard errors we directly correct for heteroskedasticity and autocorrelation.

4.3.3 The Model and Parameters

5. Constant Parameters

We assume constant parameters, the parameters are fixed unknown numbers with standard deviation larger than zero. This means that we assume that the same parameters can be used to describe the complete set of datapoints. If this is not the case it might be beneficial to define different models for different regimes of data.

6. Linear Relationship

We assume a linear relationship between the market share and the relative prices. Before the model in 4.4 can be adequately used we first test whether this linear relation suffices. For this we perform a general misspecification test for linearity. This test is known as Ramsey's Regression Equation Specification Error Test (RESET). When this test indicates linearity problems we might want to consider non-linear models. We can also try several transformations[3] on the dependent and independent variables.

4.3.4 The Probability Distribution

7. Normally Distributed Error Terms

When we assume normally distributed error terms, the OLS estimates are both consistent and (asymptotically efficient). Also the t-test for the significance of the model parameters depends on the normal distributed error terms. When the error terms are not normally distributed, the t-values are not reliable. We also assume that the error terms have a mean equal to zero. We use the Jarque-Bera test to check whether the error terms follow a normal distribution. If this is not the case, the OLS estimates are not consistent.

4.3.5 No Multicollinearity

The absence of Multicollinearity (high correlations between the regressors) is another assumption for using OLS. Multicollinearity is a statistical phenomenon in which two or more predictor variables in a OLS model or the variable and the error terms are highly correlated. In this situation the parameter estimates are not consistent. The problem is that because of the correlations we cannot adequately identify the individual effects of the predictor variables. Therefore, the elasticities estimated from the OLS model

[3]We can try log, square root, square, and inverse transformations

will become unreliable. Multicollinearity does not bias the parameter estimates, however adapting multiple variables that describe the same effect might lead to overestimation of this effect. To test for multicollinearity we compute the Variance Inflation Factor (VIF) for each of the parameter estimates. The VIF quantifies the severity of multicollinearity in the OLS model for each of these parameter estimates. The VIF is a measurement on how much the variance of the estimated parameters increases due to collinearity. A common used bound for reliable estimates is found at a VIF value of 10. Using this bound, a VIF larger than 10 may indicate a collinearity problem. To account for endogeneity a common solution is the use of instrumental variables (IV). Using the IV can be implemented using a two-stage least squares approach. In the first stage the independent variables from the original model are expressed as a linear combination of the instrumental variables. In the second stage the dependent variable is regressed on the fitted values for the independent variables from the first stage. Although the procedure for using IV is relatively easy, finding appropriate instruments is much harder.

Chapter 5

Two-Step DRPD Model

In this chapter we propose the Two-Step Dependent Region Dynamic Pricing model, referred to as DRDP model. The DRDP model is a pricing model which can be used in a tour operators dynamic pricing decisions. In addition to the IRDP model in [50] (described in section 2.2) we integrate extended consumer behavior and market information into the model. In the previous section we described how the consumer behavior can be quantified using cross elasticities. The estimated elasticities from the model in equation (4.4) are the input for the DRDP model.

First, in section 5.1 we introduce the used sets, parameters and decision variables. These are used later on in the optimization model. The cross elasticities are used to express the region demand as a function of a given pricing decision.

In section 5.2, we formulate the first step of the DRDP model. In the first step of the DRDP model, we solve the pricing problem for the regional clusters. This step can be seen as the missing link between the individual product optimization and the dependencies between the different regional clusters.

In section 5.3, we formulate the second step of the DRDP model. In the second step, the only dependency between the regions is fixed. Therefore, we can again assume independent regions in the second step of the DRDP model. Despite the fact that we use a constraint to fix the average region price, the second step of the DRDP model is very similar to the IRDP model from [50].

In section 5.4, we consider stochastic demand. We first discuss how stochastic demand is usually treated in RM models. Finally, we adjust the capacity constraints to include a buffer capacity. This buffer capacity is extra capacity that is maintained to mitigate the risk of lost demand. We introduce these constraints for uncertainty in both the first and the second step of the DRDP model.

Finally, in section 5.5 we propose a structure in which the DRDP model can be used in practice. We provide this structure to place the model into a practical perspective.

5.1 Notation

We first introduce the used sets. Subsequently, we introduce the parameters and variables.

Sets:

$i \in I$:	Set of products
$j \in J_i$:	Set of candidate prices for product i
$t \in T$:	Set of time periods
$k \in K$:	Set of resources
$S_k \subset I$:	Set of products making use of resource k
$r, o \in R$:	Set of regions
$I_r \subset I$:	Set of products corresponding to region r
$K_r \subset K$:	Set of resources corresponding to region r
$AK_r \subset K$:	Set of Accommodation Resources corresponding to region r
$FK_r \subset K$:	Set of Flight Resources corresponding to region r
$a \in A_r$:	Set of candidate prices for region r

The first 5 sets were already defined in section 2.2. We also define the set of regions R and we partition the products per region, as in section 2.2.1. We do the same thing for the capacities $K = \{K_1, K_2, \ldots, K_r\}$, we assume that a capacity can be assigned to exactly one region. K_r consists of all resources corresponding to region r, these can be accommodation or flight resources. The set of accommodation resources AK_r and the set of flight resources FK_r are defined per region. For each region r it holds that $K_r = AK_r \cup FK_r$, where $AK_r \cap FK_r = \emptyset$.

The set A_r consists of all average candidate prices for region r, this set is defined later on in equation (5.3).

Parameters:

p_{ij}:	Candidate price j for product i

c_i: Cost price of product i

q_{ijt}: Demand for product i if candidate price j is set in time period t

Cap_k: Available capacity for resource k

P_{rt}: Price of products to region r at time period t

C_r: Cost price of products to region r

M_t: Total market demand for products at time t

N_{rt}: Current market price of products to region r at time t

G_{rt}: Relative price for region r compared to the market price at time t

H_{rot}: Relative price for region r compared to other region o at time t

β_{ro}, $r = o$: Elasticity between region r and the market

β_{ro}, $r \neq o$: Cross elasticity between region r and alternative region o

Q_{rt}: Demand for products to region r at time t

The first four parameters where already defined in section 2.2. We make a clear distinction between the parameters per product (lower cases) and the parameters per regional cluster (upper cases).

Price and Cost per Region

We define the price of products to region r at time t as P_{rt}. This can be computed from the tour operator's current product prices. We do this by taking the average of all prices of products to region r at time t, see formula (5.1).

$$P_{rt} = \frac{1}{|I_r|} \sum_{i \in I_r} p_{it}, \quad \forall t \in T, \tag{5.1}$$

where $|I_r|$ is the cardinality of set I_r, which is the number of products corresponding to region r.

The average cost price of products to region r is defined as C_r, this can be computed from the tour operator's cost prices per product. We do this using formula (5.2).

$$C_r = \frac{1}{|I_r|} \sum_{i \in I_r} c_i \tag{5.2}$$

Candidate Prices per Region

After we obtained the average price of a region we can also define the set of candidate prices for each region A_r, in formula (5.3):

$$A_r = \{P_r - 2\delta\%, P_r - \delta\%, P_r, P_r + \delta\%, P_r + 2\delta\%\} \tag{5.3}$$

So we are searching for prices in the neighborhood of the current region price. The value of δ defines how wide this neighborhood is defined. The candidate price P_{rat} can be computed from P_{rt} by deviating with the percentage defined in A_r from P_{rt}.

Market Demand and Market Price per Region

M_t is defined as the total market demand for products at time t. N_{rt} is defined as the current market price of products to region r at time t.

Assumption 5.1. *We assume that the market demand and the market price are fixed over the optimization horizon (Ceteris Paribus condition). The latest known market demand is used as an estimate for the periods in the optimization horizon.*

This assumption provides us an estimate for the market demand and market prices per region. It is also possible to use estimation techniques to determine the expected market demand and market prices per time period.

Relative Prices per Region

G_{rt} is defined as the relative price for region r compared to the market price at time t, see formula (4.2). H_{rot} is defined as Relative price for region r compared to other region o at time t, see formula (4.3).

Cross Elasticities

In the regression equation in (4.4) we defined which regions have a significant cross price relation. In other words, for each region we defined which regions relative prices significantly influence the regions market share. The elasticities which are not significant in the regression are set to zero. The estimated elasticity matrix β_{ro} is defined in (5.4):

$$\beta_{ro} = \begin{cases} \widehat{\beta}_{ro}, & if\ \widehat{\beta}_{ro}\ is\ significant\ in\ (4.3) \\ 0, & otherwise \end{cases} \tag{5.4}$$

Demand for a Given Pricing Decision

Q_{rt} is the demand for region r at time t. Using the elasticities from (5.4), Q_{rt} can be expressed as a function of the pricing decision at time t. However, we first need to know the chosen prices in period t. Therefore, we first introduce decision variables Y_{rat}:

Variables:

Y_{rat}: if alternative price a is chosen for region r at time t.

Y_{rat} is a binary variable, this can be defined as in formula (5.5):

$$Y_{rat} = \begin{cases} 1, & if\ price\, a\ is\ used\ for\ region\ r\ in\ period\ t \\ 0, & otherwise \end{cases} \tag{5.5}$$

Using the cross elasticities from (5.4) and the pricing decision from (5.5), the demand function can be expressed as in (5.6):

$$Q_{rt} = M_t(\alpha_r + \sum_{a \in A_r} (\beta_{rr} G_{rat} Y_{rat} + \sum_{\substack{o \in R \\ o \neq r}} \sum_{l \in A_o} \beta_{ro} H_{raolt} Y_{rat} Y_{olt})). \tag{5.6}$$

In this equation, G_{rat} is the relative price for region r compared to the market price under pricing alternative a at time t. H_{raolt} is the relative price for region r under price alternative a compared to region o under price alternative l at time t. The market share for region r (S_{rt}) is built up from the regression equation (4.4). The first term α_r, is simply the regression constant estimated in the regression. The second term, considers all pricing possibilities for region r and other regions o. $\beta_{rr} G_{rat} Y_{rat}$ is only added to the market share if price a is chosen for region r, in this case $Y_{rat} = 1$. $\beta_{ro} H_{raolt} Y_{rat} Y_{olt}$ is only added if price a is chosen for region r and price l is chosen for other region o, in this case $Y_{rat} Y_{olt} = 1$. The market share for a given pricing decision is multiplied by the total market demand to obtain the actual demand.

Tightest Capacities per Region

Finally, for each region we want to determine the current available capacity. This is the maximum number products that still can be sold to a certain region. To calculate this, we first differentiate between flight capacities and accommodation capacities. Let $FCap_r$ be the total flight capacity for products to region r. The total flight capacity is expressed in 5.7 as the sum of all flight capacity available within this region:

$$FCap_r = \sum_{k \in FK_r} Cap_k, \quad \forall r \in R, \tag{5.7}$$

Let $ACap_r$ be the total accommodation capacity for products to region r. The total accommodation capacity is expressed in 5.8 as the sum of all accommodation capacity available within this region:

$$ACap_r = \sum_{k \in AK_r} Cap_k, \quad \forall r \in R, \tag{5.8}$$

The tightest capacity for region r is defined in (5.9) as the minimum of total flight and total accommodation capacity to every region:

$$TCap_r = min(FCap_r, HCap_r), \quad \forall r \in R, \tag{5.9}$$

5.2 DRDP STEP 1

In the first step of the DRDP model, we solve the pricing problem for the regional clusters. This step can be seen as the missing link between the individual product optimization and the dependencies between the different regional clusters. We want to determine the optimal prices for each of the regions, considering cross dependencies between the regions. We also want to make sure the expected demand for a region does not exceed the available capacity for this region. The first step of the DRDP model is a Binary Quadratic Programming (BQP) model, defined in (5.10)-(5.13).

Maximize:

$$\begin{gathered}\pi_{DRDP1} = \textstyle\sum_{r \in R} \sum_{a \in A_r} \sum_{t \in T} (P_{rat} - C_r) * \\ M_t(\alpha_r + \beta_{rr} G_{rat} + \textstyle\sum_{\substack{o \in R \\ o \neq r}} \sum_{l \in A_o} \beta_{ro} H_{raolt} Y_{rat}) Y_{olt}\end{gathered} \tag{5.10}$$

Subject to:

$$\begin{gathered}\textstyle\sum_{t \in T} M_t(\alpha_r + \sum_{a \in A_r} (\beta_{rr} G_{rat} Y_{rat} + ... \\ \textstyle\sum_{\substack{o \in R \\ o \neq r}} \sum_{l \in A_o} \beta_{ro} H_{raolt} Y_{rat} Y_{olt})) \leq TCap_r, \ \forall r \in R\end{gathered} \tag{5.11}$$

$$\sum_{a \in A_r} \sum_{t \in T} Y_{rat} = 1, \qquad \forall r \in R \tag{5.12}$$

$$Y_{rat} \in \{0, 1\}, \qquad \forall r \in R, a \in A_r, t \in T \tag{5.13}$$

When we replace $M_t(\alpha_r + \beta_{rr} G_{rat} + \beta_{ro} H_{raolt})$ by Q_{raolt}, we can simplify (5.10)-(5.13) to (5.14)-(5.17).

Maximize:

$$\pi_{DRDP1} = \sum_{r \in R} \sum_{a \in A_r} \sum_{\substack{o \in R \\ o \neq r}} \sum_{l \in A_o} \sum_{t \in T} (P_{rat} - C_r) Q_{raolt} Y_{rat} Y_{olt} \tag{5.14}$$

Subject to:

$$\sum_{t \in T} Q_{rt} \leq TCap_r, \qquad \forall r \in R \tag{5.15}$$

$$\sum_{a \in A_r} \sum_{t \in T} Y_{rat} = 1, \qquad \forall r \in R \tag{5.16}$$

$$Y_{rat} \in \{0,1\}, \quad \forall r \in R, a \in A_r, t \in T \tag{5.17}$$

In the objective, the margin $(P_{rat} - C_r)$ for a given alternative price a is multiplied with the expected demand (Q_{raolt}). This term is added when $Y_{rat} = Y_{olt} = 1$, when both price a and price l are chosen. In constraint (5.15) we make sure that for every region the expected demand is not larger than the total capacity available for this region. The total capacity for region r is defined as the minimum of total flight and total accommodation capacity to every region, see (5.9). Constraint (5.16) is a multiple choice constraint, for each of the regions one price must be chosen. Finally, constraint (5.17) states that the decision variables are binary.

Y_{rat} and Y_{olt} are the binary decision variables. Because these two binary decision variables are multiplied, the given problem is a BQP, which are usually NP-hard problems and hence practically difficult to solve (see [1]). In section 5.2.1 we describe how the quadratic objective can be linearized.

Remark 5.1. *In total there are $|R| * |A_r| * |T|$ decision variables. So the number of decision variables grows linear in the number of regions. Also the size of the problem can be regulated by changing the number of alternative prices and the number of time periods.*

5.2.1 Linearization of the BQP Model

The BQP objective in (5.14) is not linear, due to the quadratic binary term $(Y_{rat}Y_{olt})$ in the objective. The linearization in this section comes from [30]. In this section we define this exact linearization for our specific case. First, we introduce the set of ordered region combinations:

$(r,o) \in E$: Set of ordered region pairs, (with $1 \leq r \leq o \leq |R|$)

This set contains all ordered combinations of regions (r and o), in total there are $|R|^2$ combinations. Next, we introduce new variables to replace the quadratic binary terms in the objective. Thess decision variables are defined in (5.18):

$$W_{raolt} = Y_{rat}Y_{olt}. \tag{5.18}$$

Hereafter, we use the additional constraints (5.19)-(5.22) to linearize the problem into a Mixed Integer Linear Programming (MILP) problem:

$$W_{raolt} \leq Y_{rat}, \qquad \forall (r,o) \in E, a \in A_r, l \in A_o, t \in T \tag{5.19}$$

$$W_{raolt} \leq Y_{olt}, \qquad \forall (r,o) \in E, a \in A_r, l \in A_o, t \in T \tag{5.20}$$

$$W_{raolt} \geq Y_{rat} + Y_{olt} - 1, \qquad \forall (r,o) \in E, a \in A_r, l \in A_o, t \in T \tag{5.21}$$

$$W_{raolt} \geq 0, \qquad \forall (r,o) \in E, a \in A_r, l \in A_o, t \in T \tag{5.22}$$

If one of the variables Y_{rat} or Y_{oat} is 0, then also $W_{roat} = 0$. When $Y_{rat} = Y_{oat} = 1$, then $W_{roat} = 1$. This shows us that this is an exact reformulation of the original problem. We refer to [21] for a recent comparison of other possible linearization's of this problem.

Remark 5.3. *By using the linearizing equations in (5.19)-(5.22),* $4 * |E| * |A_r| * |A_o| * |T|$ *constraints are added to the problem.*

We see that the number of constrains increases significantly in case of the linearization.

In order to solve step 1 of our optimization we first rewrite the objective from (5.14) into (5.23). Since Y_{rat} is binary and W_{raolt} is a positive real number this results in a non-quadratic objective:

Maximize:

$$\pi_{DRDP1} = \sum_{r \in R} \sum_{a \in A_r} \sum_{\substack{o \in R \\ o \neq r}} \sum_{l \in A_o} \sum_{t \in T} (P_{rat} - C_r) Q_{raolt} W_{raolt} \tag{5.23}$$

We use both the constraints in (5.19)-(5.22) and the constraints in (5.15)-(5.17) to obtain the linearized MILP problem. When we solve this problem we obtain the optimal prices for each region and time period. Y^*_{rat} relates to the optimal price a for region r in time period t. The optimal prices are used in the second step of the DRDP model.

5.3 DRDP STEP 2

Like in section 2.2.1, in this step we focus on the regions independently. So the model proposed in this section is solved for each region independently. In the first step we already computed for each region which of the prices is optimal for each of the regional clusters Y^*_{rat}. Therefore, the optimal average price per regional cluster per time period is defined in (5.24):

$$P^*_{rt} = \sum_{a \in A_r} P_{rat} Y^*_{rat} \tag{5.24}$$

Remark 5.4. *It is important to note that price per region determined in 5.24 is fixed during the second step of the optimization. Therefore, the only dependency between the regions is fixed. This means that we can again assume independent regions in the second step of the DRDP model.*

From equation (5.24) we know which price is optimal in which time period. Therefore, we can simply add the constraint defined in (5.25) to the model from section 2.2.

$$\frac{1}{|I|} \sum_{i \in I} \sum_{j \in J_i} p_{ij} x_{ijt} = P^*_{rt}, \qquad \forall t \in T \tag{5.25}$$

This constraint makes sure that for every time period the average price of the chosen prices is equal to the optimal average price found in step 1 of DRDP model.

When we add constraint (5.25) to the model from section 2.2, we obtain the second step of the DRDP model. The goal of the second step of the DRDP model is to maximize the expected revenue over the defined time period T. In the pricing decision we have to decide which price to choose for each product in each time interval. To do this, we again use decision variables $x_{ijt} \in [0, 1]$. x_{ijt} is defined as the fraction of time interval t price j is chosen for product i. The price fractions form a price path for every time interval. The resulting price path is used such that the price path is decreasing (i.e. the highest price with a non-zero price fraction is chosen first and the lowest price with a non-zero price fraction is chosen last). The optimal price fractions are determined using the Linear Programming (LP) model in (5.26)-(5.30).

Maximize:

$$\pi_{DRDP2} = \sum_{i \in I} \sum_{j \in J_i} \sum_{t \in T} (p_{ij} - c_i) q_{ijt} x_{ijt} \tag{5.26}$$

Subject to:

$$\sum_{i \in S_k} \sum_{j \in J_i} \sum_{t \in T} q_{ijt} x_{ijt} \leq Cap_k, \quad \forall k \in K \tag{5.27}$$

$$\frac{1}{|I|}\sum_{i\in I}\sum_{j\in J_i} p_{ij}x_{ijt} = P^*_{rt}, \qquad \forall t \in T \tag{5.28}$$

$$\sum_{j\in J_i} x_{ijt} = 1, \qquad \forall i \in I, t \in T \tag{5.29}$$

$$0 \le x_{ijt}, \qquad \forall i \in I, j \in J, t \in T \tag{5.30}$$

The objective from (5.26) sums over all periods, products and possible candidate prices. For each period, the revenue for each product $(p_{ij} - c_i)$ is multiplied by the expected demand for this price path $(q_{ijt}x_{ijt})$. The objective of this model is to choose the price path for each product that maximizes the total expected revenue over the given time horizon T.

Using constraint (5.27), we make sure that for each of the resources $k \in K$ the demand does not exceed the available capacity of this resource. Constraint 5.28 makes sure we use the optimal price within each regional cluster. This is the only difference in this step with the IRDP model formulated in (2.1)-(2.4). Constraints (5.29) and (5.30) are constraints on the price path. These make sure that the price path meets several conditions. Constraint (5.29) ensures that for each time period the price path is completely defined, the fractions must sum up to one for each time period t. Additionally, in constraint (5.30) we make sure that all fractions are non-negative.

Remark 5.5. *It is important to note that we fixed the price per region by using the constraints in (5.28), also see remark 5.4. Therefore, optimality from step 1 is never violated in step 2 of the DRDP model. This implicates that the eventual pricing decision is always optimal in terms of the regional clusters.*

5.4 Stochastic Demand

The first and second step DRDP models described in this chapter are deterministic. The demand is treated as given using point estimates. In constraints (5.15) and (5.27), we make sure that the expected demand is smaller than the capacity. In (5.15) this is done for each of the regions r and in (5.27) this is done for each of the resources k. Because we treat the expected demand as given, the pricing decision from this model is only optimal if the realized demand is equal to the expected demand.

In reality the demand is much more uncertain. For the airline industry, [7] proposed a discretization approach to integrate possible scenario outcomes with given probabilities of these scenarios into the objective function. However, [7] and [51] find that the use of such a stochastic model in the airline industry does not necessarily improve the results. Moreover, [7] concludes that stochastic booking limit models do not improve the results because the stochastic nature of demand is already included in deterministic booking limit models.

Alternatively, [46] proposed a simulation based 'stochastic' model. The discrete optimization model is used to solve the model with simulated demand as expected demand. After multiple optimizations, the gradient can be estimated as the average of all solutions. This simulation approach offers a lot of flexibility in modeling stochastic demand for various distributions, however a disadvantage is that the model has to be solved multiple times.

In a case of hotel RM, [40] compared both the stochastic and simulation approach concluding that the simulation based model outperformed the actual stochastic models. Also in this experiment the disadvantage of high computation times for the simulation model is raised. In contrary, according to [40] stochastic models can improve the results of the bid-price model.

By ways of experiment, we propose a completely different approach. Our approach of treating uncertain demand is commonly used in inventory management (see for example, [2]). In inventory management, the inventory for a given product along with the demand distribution are often used to calculate the probabilities of stock outs. These probabilities are then used to determine a safety stock, an extra stock that is maintained to mitigate risk of stock outs. In our case, there are resources with uncertain demand and limited capacities (inventory). In the remainder of this section we show how we can use the central limit theorem to calculate the demand probabilities. We show how the constraint in (5.15) can be adjusted to include for demand uncertainty per region. Hereafter, we show how the constraint in (5.27) can be adjusted to include for demand uncertainty per resource.

5.4.1 Stochastic Demand per Region

Let θ_r be the probability that the demand for region r within time horizon T exceeds the total capacity for this region ($TCap_r$).

$$\theta_r = P(Q_r > TCap_r) \tag{5.31}$$

with $Q_r = \sum_{t \in T} Q_{rt}$, the demand for region r over time horizon T. Using the central limit theorem we can assume that $Q_r \sim N(\mu_r, \sigma_r)$, with μ_r and σ_r the mean and standard deviation of the demand for region r, respectively. We indirectly assume that the total region demand is built up from the individual product demands, and that none of these product demands exhibits dominant behaviour (i.e. none of the individual product demands is excessively larger than others). We can use standard normal pdf function $\phi()$ to calculate the demand probabilities. To make sure that the demand for r does not exceed the capacity with a certain probability θ_r, we can rewrite constraint (5.15) into 5.32:

$$\sum_{t \in T} Q_{rt} + \phi^{-1}(1-\theta_r)\sigma_r \leq TCap_r, \quad \forall r \in R, \tag{5.32}$$

with Q_{rt} the demand as expressed in equation 5.6. Where $\phi^{-1}(1-\theta_r)\sigma_r$ is called the safety capacity, the extra capacity that is maintained to fulfill the probability of lost demand θ_r.

5.4.2 Stochastic Demand per Resource

Let ζ_k be the probability that the demand for resource k exceeds the capacity for this resource (Cap_k).

$$\zeta_k = P(q_k > Cap_k), \tag{5.33}$$

with $q_k = \sum_{i \in S_k} \sum_{j \in J_i} q_{ij} x_{ij}$, the total demand for resource k over time horizon T. Using the central limit theorem we can assume that $q_k \sim N(\mu_k, \sigma_k)$, with μ_k and σ_k the mean and standard deviation of the demand for resource k, respectively. We indirectly assume that the total resource demand is built up from the individual product demands, and that none of these product demands exhibits dominant behaviour (i.e. none of the individual product demands is excessively larger than others). We can use standard normal pdf function $\phi()$ to calculate the demand probabilities. To make sure that the demand for k does not exceed the capacity with a certain probability ζ_k, we can rewrite constraint 5.27 into 5.34:

$$\sum_{i \in S_k} \sum_{j \in J_i} \sum_{t \in T} q_{ijt} x_{ijt} + \phi^{-1}(1-\zeta_k)\sigma_k \leq Cap_k, \quad \forall k \in K \tag{5.34}$$

Where $\phi^{-1}(1-\zeta_k)\sigma_k$ is called the safety capacity, the extra capacity that is maintained to fulfill the probability of lost demand ζ_k. Using the constraints in (5.32) and (5.34) instead of the original constraints, we make sure that the demand does not exceeds capacity for a certain probability. Note that for θ_r and ζ_k equal to 0.5 we obtain $\phi^{-1}(0.5) = 0$. In this case we have the original constraints.

The advantage of stochastic constraints is that the user is more flexible. By defining the probability that the demand exceeds capacity, the tour operator is able to control the expected lost demand. This can be used to fulfill different strategies for different optimization periods. A strategy could be to gain market share for a certain period, by choosing a smaller probability of lost demand. The disadvantage of stochastic constraints like this is that we need an estimate for the standard deviation of demand. The computational experiments considering the stochastic constraints can be found in section 7.5.6.

5.5 Intermezzo: Using the Model in Practice

This section proposes a structure in which the DRDP model can be used in practice. We provide this structure to place the model into a practical perspective. The flow diagram in figure 5.1 depicts how the DRDP model fits in the process of an online tour operator.

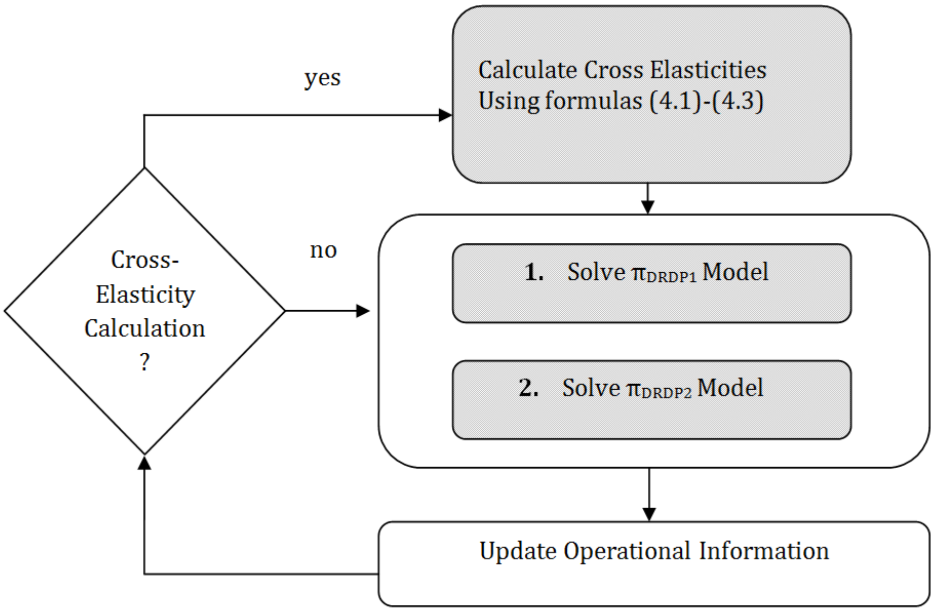

Figure 5.1: Two-Step Algorithm

To make sure the model adapt to changing environment we make use of a rolling horizon of several years of data. This horizon of data is called the Operational Information. Because the travel market changes rapidly it is a tradeoff between less 'up to date' data and more 'possibly outdated' data. Every day, the horizon 'rolls' one day further. The complete horizon of data is used to create demand forecasts and to calculate the elasticities. To keep the elasticities up to date it is required to estimate the elasticities frequently. The data from which the elasticities are estimated is aggregated per week. Therefore, the elasticities are updated once a week.

Subsequently, the updated elasticities are used to solve the DRDP model in the following week. Each night, the operational information of that day is gathered and updated. The bookings and cancellations are processed and the remaining capacities are determined. After the operation information is updated, the DRDP model is solved and the optimized prices are stored in the database. Hereafter, the optimized prices are uploaded to the website.

Part III

Results

Chapter 6

Analytical Comparison IRDP and DRDP model

In this chapter we make a analytical comparison between the DRDP and IRDP model.

In section 6.1, we focus on the number of price combinations per time period in each of the models. Among these price combinations, the model has to choose the most profitable pricing decision. The number of price combinations is used as a benchmark for the complexity of the model. In the DRDP model we use a two-step approach. In the second step we fixed the optimal cross regional effects. Because of this, we can assume independent regions in the second step of the DRDP model. In the first section we investigate the implication of this approach on the problem complexity. We first compare the DRDP model to a hypothetical model in which we do not use a two-step approach, we refer to this model as the Complete One-Step (COS) model. Next, we compare the DRDP model to the IRDP model. We also provide an example in which we show the magnitude of the differences in complexity between the models.

In section 6.2, we compare the objectives of the IRDP and the DRDP model. We show that the objective of the models does not per definition represent the expected profit.

6.1 Number of Price Combinations

In the IRDP model in equations (2.1)-(2.4), the expected number of purchasing customers for a product is determined from the historical price demand relationship. In this case, the estimated demand for a product is considered a function of the chosen price of this product. The products within a region are dependent because they share the same resources but also because they share a pool of potential customers. When we include cross elasticites, the relationships between the products become more complicated. Moreover, when the regions become dependent, we can no longer decompose the problem into independent sub problems per region (as we described in section 2.2.1). In figure 6.1 we see that the pricing (LP) models which were assumed to be independent now influence each other. The pricing decision of a region now influences the demand of all other (competing) regions. Instead of $|R|$ independent models we now have one large pricing model, referred to as the complete one-step model. Although we did not formulate this model, we use the number of price combinations in this 'hypothetical' model as a benchmark.

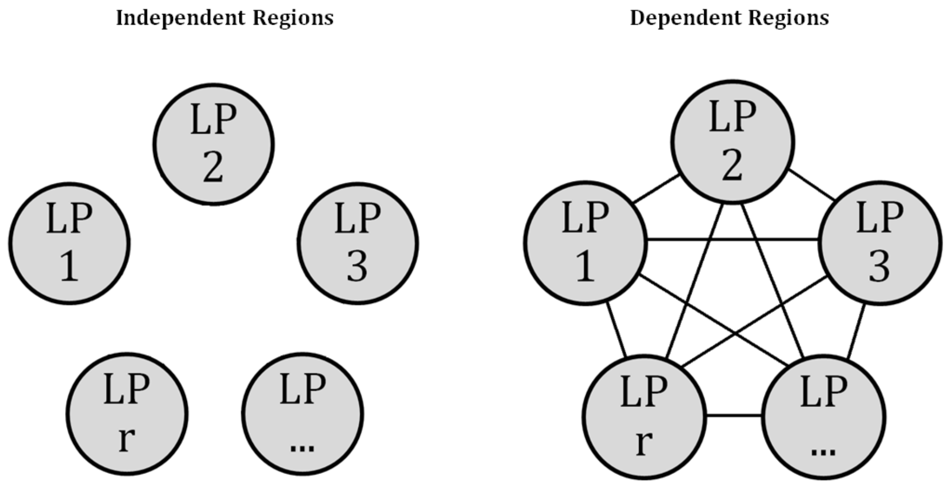

Figure 6.1: Structure of the Problem: Independent vs Dependent Regions

The complexity of the problem mainly depends on the involved number of regions, products, price alternatives and time periods. In the IRDP model, we considered $|R|$ regions with $|I_r|$ products per region, for each product we considered $|J_i|$ possible prices. This leads to $|J_i|^{|I_r|}$ pricing combinations per region per time period. In the case of independent regions, there are $|R||J_i|^{|I_r|}$ pricing combinations. In the case of dependent regions, the amount of pricing combinations per period becomes $|J_i|^{|R||I_r|}$. In this case, we have one big model in which all products are dependent. In the DRDP model we first solve the pricing problem on the region level with $|J_i|^{|R|}$ possible price combinations. Subsequently, we solve the second step of the model for each region independently (like the IRDP model $|R||J_i|^{|I_r|}$ price combinations). In total, the DRDP model contains $|J_i|^{|R|} + |R||J_i|^{|I_r|}$ price combinations ($|J_i|^{|R|}$ for the first step and $|R||J_i|^{|I_r|}$ for the second step).

6.1.1 DRDP model vs Complete One-Step model

First of all, we are interested in the number of price combinations in the two-step DRDP approach compared to the complete one-step model. In theorem 6.1 we state that in case of dependent regions, the number of price combinations in the DRDP model is strictly smaller than the number of price combinations in the complete one-step model.

Theorem 6.1.

$$|J_i|^{|R|} + |R||J_i|^{|I_r|} < |J_i|^{|R||I_r|}, \forall(|R|, |I_r|, |J_i|) \in \mathbb{Z}, (|R|, |I_r|, |J_i|) > 1 \quad (6.1)$$

Proof:

Let $x = |R|$, $y = |I_r|$ and $z = |J_i|$. We want to prove that $z^x + xz^y < z^{xy}$ with ($x, y, z \in \mathbb{Z}$) and ($x, y, z > 1$). To do this, we use three-dimensional mathematical induction. Let, $P(m, n, k)$ be the inequality involving arbitrary variables m,n and k. Mathematical induction states that if all of the following conditions hold, we have proven that the inequality holds:

C1. $P(2, 2, 2)$ *holds*

C2. $P(m + 1, n, k)$ *holds, given that* $P(m, n, k)$ *holds*

C3. $P(m, n + 1, k)$ *holds, given that* $P(m, n, k)$ *holds*

C4. $P(m, n, k + 1)$ *holds, given that* $P(m, n, k)$ *holds*

With ($m, n, k \in \mathbb{Z}$) and ($m, n, k > 1$)

C1.

We want to show that $P(2, 2, 2)$ holds. We can simply plug in the values in the inequality to obtain: $2^2 + 2 * 2^2 = 12$ and $2^{2*2} = 16$ to show that C1. holds.

For conditions C2.-C4. we use the $P(m, n, k)$ assumption, we assume that the following inequality holds:

A1. $k^m + mk^n < k^{mn}$, *for arbitrary ($m, n, k \in \mathbb{Z}$) and ($m, n, k > 1$)*

C2.

We want to show that $k^{(m+1)} + (m + 1)k^n < k^{(m+1)n}$ holds, given A1.: We rewrite the right hand side of the inequality into $k^{mn} * k^n$, and divide both sides by k^n to obtain:

$$k^{m+1-n} + (m + 1) < k^{mn}$$

Now we look at assumption A1. Since $m > 1$, $n > 1$ and $k > 1$ we know that $k^{m+1-n} < k^m$ and that $(m + 1) < mk^n$. Therefore using A1. we get:

$$k^{m+1-n} + (m + 1) < k^m + mk^n < k^{mn}$$

Which shows C2. holds for arbitrary ($m, n, k \in \mathbb{Z}$) and ($m, n, k > 1$)

C3.
We want to show that $k^m + mk^{(n+1)} < k^{m(n+1)}$ *holds, given A1.: We rewrite the right hand side of the inequality into* $k^{mn} * k^m$*, and divide both sides by* k^m *to obtain:*

$$1 + mk^{n+1-m} < k^{mn}$$

Now we look at assumption A1. Since $m > 1$*,* $n > 1$ *and* $k > 1$ *we know that* $1 < k^m$ *and that* $mk^{n+1-m} < mk^n$*. Therefore using A1. we get:*

$$1 + mk^{n+1-m} < k^m + mk^n < k^{mn}$$

Which shows C3. holds for arbitrary $(m, n, k \in \mathbb{Z})$ *and* $(m, n, k > 1)$*.*

To prove C4., let us first rewrite A1. into $\frac{k^m}{k^{mn}} + \frac{mk^n}{k^{mn}} < 1$ *by dividing both sides of the inequality by* k^{mn}*.*

C4.
We want to show that $(k+1)^m + m(k+1)^n < (k+1)^{mn}$ *holds, given A1.: Divide both sides of the inequality by* $(k+1)^{mn}$*, we obtain:*

$$\frac{(k+1)^m}{(k+1)^{mn}} + \frac{m(k+1)^n}{(k+1)^{mn}} < 1$$

Now we look at the rewritten assumption A1. Since $m > 1$*,* $n > 1$ *and* $k > 1$ *we know that* $\frac{(k+1)^m}{(k+1)^{mn}} < \frac{k^m}{k^{mn}}$ *and that* $\frac{m(k+1)^n}{(k+1)^{mn}} < \frac{mk^n}{k^{mn}}$*. Therefore using A1. we get:*

$$\frac{(k+1)^m}{(k+1)^{mn}} + \frac{m(k+1)^n}{(k+1)^{mn}} < \frac{k^m}{k^{mn}} + \frac{mk^n}{k^{mn}} < 1$$

Which shows C4. holds for arbitrary $(m, n, k \in \mathbb{Z})$ *and* $(m, n, k > 1)$*.*

This concludes the proof that $z^x + xz^y < z^{xy}$ *with* $(x, y, z \in \mathbb{Z})$ *and* $(x, y, z > 1)$*.* □

Using theorem 6.1 we know that in case of dependent regions, the DRDP approach reduces the number of pricing combinations per definition.

6.1.2 DRDP model vs IRDP model

We are interested in the comparison between the number of price combinations in the DRDP model compared to the IRDP model. In theorem 6.2, we state that the number of price combinations under the IRDP model is strictly smaller than the number of price combinations under the DRDP model.

Theorem 6.2.

$$|R||J_i|^{|I_r|} < |J_i|^{|R|} + |R||J_i|^{|I_r|}, \forall (|R|, |I_r|, |J_i|) \in \mathbb{Z}, (|R|, |I_r|, |J_i|) > 1 \quad (6.2)$$

Proof: *We know that* $|J_i| > 0$ *and* $|R| > 1$, *hence* $|J_i|^{|R|} > 1$. *Therefore,* $|R||J_i|^{|I_r|} < |J_i|^{|R|} + |R||J_i|^{|I_r|}$ $\square$

Using theorem 6.2, we know that the number of price combinations in the IRDP model is strictly smaller than the number of price combinations in the DRDP model. Therefore, the complexity grows by using the DRDP model instead of the IRDP model. The percentage growth in the number of price combinations is derived in the next section.

% Growth in Price Combinations

We are mainly interested in the complexity of the DRDP model compared to the IRDP model. Per theorem 6.2, we know that the number of price combinations for the DRDP model is higher than the number of price combinations in the IRDP model. In theorem 6.3, we define the percentage growth in price combinations as a result of using the DRDP model instead of the IRDP model.

Theorem 6.3. *For an arbitrary problem instance, using the DRDP model instead of the IRDP model leads to an additional* $\frac{|J_i|^{|R|-|I_r|}}{|R|}\%$ *in price combinations.*

Proof: *Let* $x = |R|$, $y = |I_r|$ *and* $z = |J_i|$. *We assume that* $(x, y, z) \in \mathbb{Z}$ *and* $(x, y, z) > 1$. *The percentage growth in price combinations of the DRDP model compared to the IRDP model can be formulated as:*

$$\Delta\% = \frac{(z^x + xz^y) - xz^y}{xz^y}$$

This can be reformulated as:

$$\Delta\% = \frac{z^{x-y}}{x} \quad \square$$

In reality, this percentage will be relatively small because the number of products per region is expected to be much larger than the number of regions.

In theorem 6.3, we expressed the percentage growth in price combinations as a result of using the DRDP model instead of the IRDP model. This way we have quantified the growth in price combinations. Using theorem 6.3, we know that for an arbitrary problem instance with $x = |R|$ regions, $y = |I_r|$ products per region, and $z = |J_i|$ alternative prices per product the number of price combinations will grow with $\frac{z^{x-y}}{x}\%$.

A relatively high percentage growth will be realized when the number of regions is relatively high compared to the number of products pre region. In reality, the number of products per region will not be smaller than the number of regions.

6.1.3 Magnitude of Price Combination Difference

We compared the DRDP approach with the IRDP and the complete one-step approach. For each approach, we are interested in the growth of the number of pricing combinations for a given number of regions, products per region and number of price alternatives. We want to know the magnitude of these differences in pricing combinations for different set sizes. The following example provides us this insight:

Example 6.1: Comparison Different Approaches. *We look at $x = |R|$ regions with $y = |I_r|$ products per region r and $z = |J_i|$ alternative prices per product i. In the case of the IRDP model, we have xz^y possible price combinations. In case of Complete One-Step (COS) model we have z^{xy} price combinations. In case of our Two-Step DRDP model we have $z^x + xz^y$ price. In table 6.1 we see how the number of price combinations evolves for different set sizes.*

			Independent Regions IRDP Model	Dependent Regions COS model	Dependent Regions DRDP model
$x = \|R\|$	$y = \|I_r\|$	$z = \|J_i\|$	xz^y	z^{xy}	$z^x + xz^y$
2	5	5	$6.25 * 10^3$	$9.77 * 10^6$	$6.28 * 10^3$
4	5	5	$1.25 * 10^4$	$9.54 * 10^{13}$	$1.31 * 10^4$
6	10	5	$5.86 * 10^7$	$8.67 * 10^{41}$	$5.86 * 10^7$
8	10	5	$7.81 * 10^7$	$8.27 * 10^{55}$	$7.85 * 10^7$
10	20	5	$9.53 * 10^{14}$	$6.22 * 10^{139}$	$9.53 * 10^{14}$
12	20	5	$1.14 * 10^{15}$	$5.66 * 10^{167}$	$1.14 * 10^{15}$

Table 6.1: Growth of Price Combinations

We must note that the results from table 6.1 hold for each time period. Example 6.1 clearly illustrates the large benefits of the DRDP model in terms of the number of price combinations. The number of alternative prices is assumed fixed at 5 price alternatives. The number of regions is varied from 2 to 12 and the number of products per region is varied from 5 to 20. In reality, the number of regions is not expected to grow very fast. Due to a large number of possible combinations, the number of products per region can grow fast. In case we do not use the DRDP approach the number of price combinations grows excessively. In case of 12 regions, 20 products per region and 5 price alternatives per product, the number of price combinations is $5.66 * 10^{167}$. When the DRDP approach is used the number of price combinations is reduced to $1.14 * 10^{15}$ combinations. So, by dividing the pricing decision into two steps, the DRDP approach is able to reduce the number of price combinations considerably.

The difference between the IRDP and DRDP approach is only $|J_i|^{|R|}$ price combinations (this is 5^{12} for the large example). So compared to the IRDP approach, the DRDP approach does not strongly increase the number of price combinations, see theorem 6.3. In the case of 12 regions, 20 products per region and 5 price alternatives per product, the growth percentage is approximately 0.00%.

6.2 Objectives and Expected Profit

An important thing to notice is that the objective from the IRDP model (given in (2.1), denoted by Π_{IRDP}) and the objective from step 2 of the DRDP model (given in (5.26), denoted by Π_{DRDP2}) are represented by the same formula. Therefore, one might be tempted to compare these objectives in order the compare the expected profit for both models. However, the objectives Π_{IRDP} and Π_{DRDP2} do not represent the actual expected profit. In the next section we describe how the actual profit can be calculated. We next show that the optimal value from the IRDP model (Π^*_{IRDP}) will be larger or equal than the optimal value from the second step of the DRDP model (Π^*_{DRDP2}).

Theorem 6.4.

$$\Pi^*_{IRDP} \geq \Pi^*_{DRDP2} \tag{6.3}$$

Proof:
When we compare the IRDP model (given in (2.1)-(2.4)) to the second step of the DRDP model (given in (5.24)-(5.28)) we see that these mathematical programs are almost identical. Let $\mathcal{S}_{IRDP}$ be the feasible region defined in (2.2)-(2.4). Let $\mathcal{S}_{DRDP2}$ be the feasible region defined in (5.25)-(5.28). Because (5.25)-(5.28) contains all constraint from (2.2)-(2.4) and more constraints we know that $\mathcal{S}_{DRDP2} \subseteq \mathcal{S}_{IRDP}$. Because the problems also have an identical objective we know that $\Pi^*_{IRDP} \geq \Pi^*_{DRDP2}$. □

6.2.1 Calculate Expected Profit from Π_{IRDP} and Π_{DRDP2}

We proved that $\Pi^*_{IRDP} \geq \Pi^*_{DRDP2}$, so when we compare these objectives, the IRDP objective will always be the highest. The reason for this is that in the objectives π_{IRDP} and π_{DRDP2} the cross regional profits are not incorporated. Because of this, we have to adjust the objectives to obtain the correct expected profit including the expected cross regional profit. We can use the π_{DRDP1} objective (given in (5.14)) to correct for the expected cross profit generated by competing regions pricing decisions. We can correct the objectives (Π_{IRDP} and Π_{DRDP2}) for the cross regional profit in order to obtain the total expected profit.

Additional Cross Regional Profit

In the IRDP model, for each of the products, a price elasticity of demand is defined. Choosing for an alternative price of a product leads to a change in demand for that product. This change in demand for the individual products is calculated using this price elasticity. In the second step of the DRDP model, the products also have their own price elasticities. Using only price elasticities of demand, the cross regional effects are ignored in this step. Note that in case of the DRDP model this does not have any negative consequences on the pricing decision itself, see remark 5.5. In step 2 of the DRDP model, the average price for a region is fixed to the optimal region price from step 1. A consequence is that the cross regional profit is just a constant term that can be calculated after the regional pricing decision in step 1 of the DRDP model.

Let P_{rt} be the base price for region r at time t calculated from the base product prices p_{it}, using formula (5.1). Let P^*_{rt} be the optimal region prices obtained after step 1 of the DRDP model.

The relative price changes after step 1 of the DRDP model, for time period t, are defined in (6.4):

$$\Delta H_{rot} = \frac{P^*_{rt}}{P^*_{ot}} - \frac{P_{rt}}{P_{ot}}. \tag{6.4}$$

When the relative price of region r compared to other region o has increased, we assume that region r loses customers to region o. However, if the price increase is sufficient this might be profitable. The total cross profit for region r (χ_r) over time horizon T is defined in (6.5):

$$\chi_r = \sum_{\substack{o\in R\\ o\neq r}} \sum_{t\in T} (P^*_{rt} - C_r) M_t (\beta_{ro} \Delta H_{rot}). \tag{6.5}$$

This formula uses the elasticities from the regression in (4.4), the change in relative price results in a change in market share ($\beta_{ro}\Delta H_{rot}$). This is multiplied with the total market demand (M_t) in order to compute the change in actual customers. The change in customers is multiplied with the margin of the region for a given pricing decision ($P^*_{rt} - C_r$).

Note that the cross profit can also be calculated for the IRDP model. After a pricing decision, we first have to calculate the region price and cost from the optimal product prices (using formulas (5.1) and (5.2)).

The total cross profit summed over all regions can be added to the objectives from the first step of the DRDP model and the IRDP model. In these objectives, the cross regional profit was not incorporated. For the DRDP model, the total expected profit (Ψ_{DRDP}) including the cross regional profit ($\sum_r \chi_r$) can be calculated using the formula in (6.6):

$$\Psi_{DRDP} = \pi_{DRDP2} + \sum_r \chi_r. \tag{6.6}$$

In this equation, π_{DRDP2} represents the regions 'own' profits whereas $\sum_r \chi_r$ represents the regions 'cross profits'. Note that in the DRDP model, the cross profits are optimal, see remark 5.5. Therefore, the cross profits from step 2 of the DRDP model are per definition higher or equal to the cross profits from the IRDP model.

Chapter 7

Case Study

In this chapter we provide a case study of a large tour operator called Sunweb. In the section 7.1 we provide some background information.

In section 7.2, we describe the available data, we focus on 4 main destinations of Sunweb. These destinations account for approximately 90% of the total bookings.

Before the elasticities can be measured using regression, we have to make sure that the regression parameters can be trusted. Therefore, in section 7.3, we perform a pre-modeling data analysis. We first look at the strategy of focusing on the relative price, market share relationship. Next we test the market share series on stationarity. We also describe the used outlier detection mechanism. In the travel industry, people are influenced by many factors other than price. Marketing activities can cause large demand peaks that cannot be explained by price changes. To be able to determine the effects of price changes on demand, outlier detection is very important.

In section 7.4 we describe the regression analysis proposed in section 4.1.2. We estimate the elasticities from the actual booking data and we perform several statistical tests to guarantee the validity of the estimates. We also used the simulation technique described in section 4.2 to estimate the elasticities of demand from the estimated cross elasticities.

In section 7.5 we use simulation to compare the results of different pricing models under different circumstances. We define multiple test cases, representing different possible situations. We vary three different attributes of a test case: capacity, margins and demand volatility. The aim is that the test cases provide a realistic mix of different possible situations. First, we compare the models after the first step of the DRDP model. Hereafter, we compare the models after the second step of the DRDP model. To be able to solve the second step of the model, we first have too extend some of the test cases. We are interested in the average profit generated for each of the models under different circumstances. We are also interested in the robustness of the models. The simulation outcomes provide us insight in the behavior of the optimization models, they also show the possible value of the DRDP model. We also discuss the results for the DRDP model using the stochastic constraints described in section 5.4. These stochastic constraints make use of a buffer capacity, extra capacities that are maintained to mitigate risk of lost demand. This approach of treating uncertain demand is commonly used in inventory management (see for example, [2]). We are interested in the effect of these constraints on both the profit and the lost demand.

In the last section 7.6, we consider the realized computation times of both the DRPD and IRDP model. In section 6.1, we already analyzed the problem complexity of the different models. We expect that the computation times are closely related to the complexity of the models.

7.1 Background

7.1.1 Sunweb

In this study we will focus on Sunweb, part of the Sundio Group[1]. Currently, Sundio is one of the largest tour operators and webshops in the Netherlands. Since 2007 the Sundio Group has been the umbrella organization corresponding to 14 different brand names. Sunweb is the largest of these brands, with a substantial share of the total Sundio bookings. Sunweb has been operating as a direct sales organization for flight package holidays to sunny destinations predominantly within Europe since 2000. They mainly provide summer holidays (Sunweb-Summer) but also Winter holidays (Sunweb-Winter). Figure 7.1 depicts the travel countries that Sunweb-Summer offers.

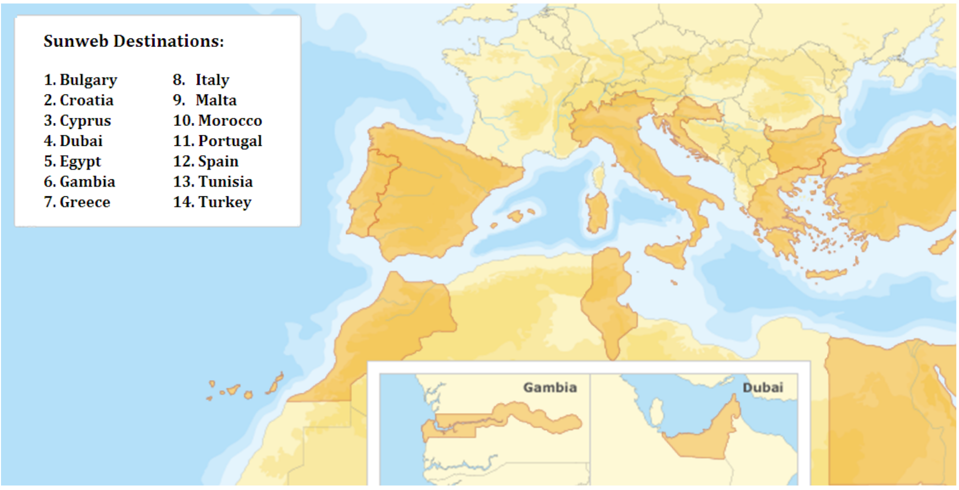

Figure 7.1: Holiday Countries of Sunweb-Summer

[1]for more information on Sundio Group or Sunweb see: http://www.sundiogroup.com

In this picture we see that Sunweb-Summer is mainly active in Western-Europe, but also in Africa, and Dubai. In the remainder of this case study, we focus on the booking data of Sunweb-Summer. This data is described in section 7.2. We first provide some background information on GfK Retail and Technology.

7.1.2 GfK Retail and Technology

GfK Retail and Technology[2] is the global leader in sales reporting and market intelligence for technical consumer goods markets. GfK Retail and Technology is active in various markets, including travel and tourism. The main clients of GfK Travel and Tourism include Tour Operators, Airports, Travel Insurers, Rental Car Companies and Hotels.

GfK Retail and Technology works in direct partnership with travel agencies and tour operators in Germany, United Kingdom, France, Italy, the Netherlands and Russia. The added value of GfK is that they provide unique market intelligence based upon live booking information. This booking information helps tour operators to get a better understanding of the highly competitive market. Besides this, the market information can be used to support tactical planning decisions, pricing decisions and marketing campaigns.

7.2 Data Description

In this study we focus on Sunweb's most important countries: Greece, Spain, Turkey and Portugal. In figure 7.2 we see the number of bookings for each country as a percentage of the total bookings. Spain and Greece are the main contributors to the total number of bookings. With approximately 90% of the total Sunweb bookings these four countries provide a good representation of the total booking process within Sunweb.

The dataset used in this chapter consists of the booking data from the first week of 2008 until the 24th week of 2011. The data is aggregated as the total number of bookings per week and the corresponding average price of these bookings. In total, the dataset consists of 181 aggregated weeks. For the same period we possess the number of market bookings and the average market price per region per week. The Market data is provided by GfK, an organization that delivers market information (described in section 7.1.2). We use the formulas (4.1) and (4.2) to compute the market share and the relative price of Sunweb compared to the rest of the market. We are only interested in the rest of the market that is able to influence the number of bookings of Sunweb. Therefore, the market is defined as: tour operators

[2]for more information on GfK Retail and Technology see: http://www.gfkrt.com/

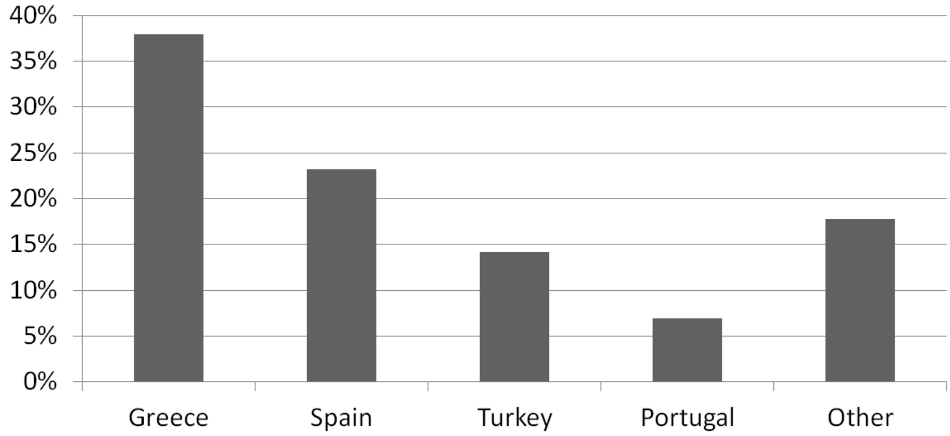

Figure 7.2: Sunweb bookings per country as a percentage of the total bookings

operating in the Netherlands (according to definition 4.1). Incorporating other travel organizations will only disturb the expected relationships.

7.3 Pre-Modeling Data Analysis

An accurate pre-modeling data analysis is necessary to make sure that we can correctly interpret the regression results. We first show that looking at the price-demand relationship can cause trouble because the data is highly seasonal. We also show that correcting for the market bookings and the market price is a successful strategy. Next, we test the market share data for stationarity. Hereafter we describe how possible outliers are treated. We also describe why this outlier detection is important in this analysis.

7.3.1 Actual Prices vs. Relative Prices

Usually the relationship between price and demand is a negative one. The coefficient in this relationship is often used as the price elasticity of demand. For tour operators it is often not possible to estimate price elasticities this way. Figure 7.3 depicts a scatterplot between price and demand per week for Turkey.

When we fit a regression line to the scatter plot in figure 7.3, we obtain a positive (not significant) relationship with a R^2 of 0.01. This positive relationship is caused by a seasonal demand and pricing pattern. We describe two periods that cause a positive relationship between price and demand. This phenomenon was already described in example 4.2.

First, we consider the period January-February: most people book their summer holiday in this period. The selling season also starts in January

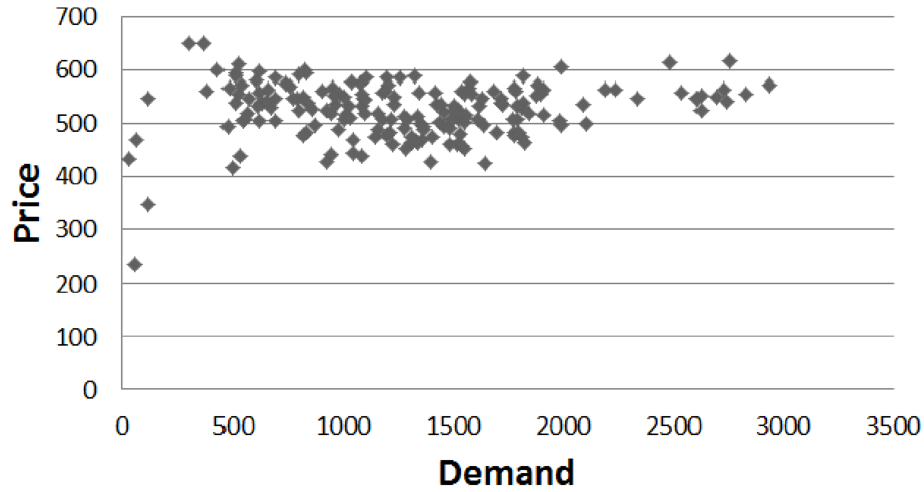

Figure 7.3: Turkey, Price(P_{Turkey}) vs. Demand(Q_{Turkey})

resulting in high prices. The dots with high price and high demand in figure 7.3 typically represent the weeks in this period.

Second, we look at the summer holiday in June-August: less people book their summer holiday in this period. However, the people that do are willing to accept the fact that there is less certainty and less choice to get lower prices. In figure 7.3 we clearly see the low price low demand dots. The very low prices (around 200 and 300) with low demands are typical examples of 'last-minute' weeks.

Both of these periods influence the relationship between price and demand to become positive. Therefore, we are not able to estimate the price elasticities using a linear regression (without seasonal dummies).

In section 4.1.1 we described how we avoid this problem using the relative price in a given week compared to the market share in this week. In this approach we assume that for each region the market share is more or less constant over time. Figure 7.4 depicts the relationship between the relative price of products to Turkey and the corresponding market share.

When we fit a regression line to this data, a negative and significant relationship is found with a R^2 of 0.37. Therefore, we can conclude that the relative price of Sunweb's products to Turkey significantly influences the market share of Sunweb to Turkey. For the other regions we find similar relationships.

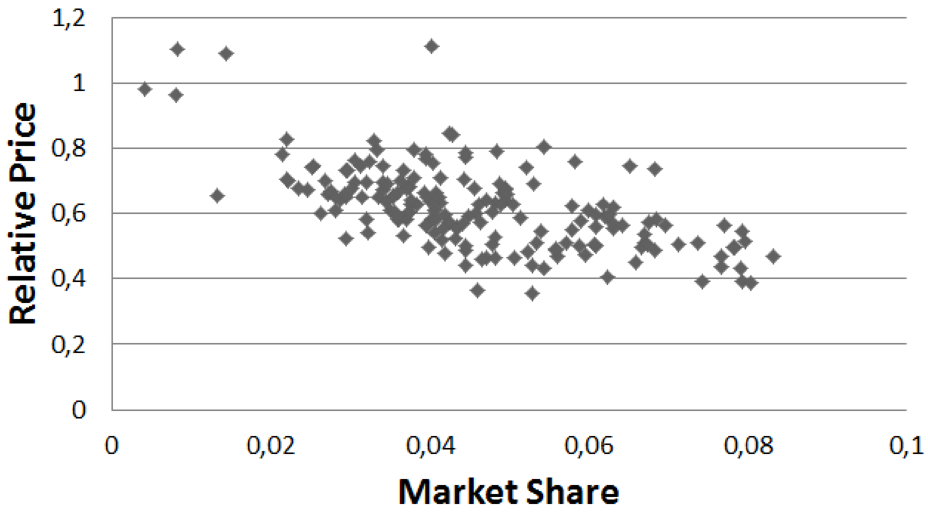

Figure 7.4: Turkey, Relative Price(G_{Turkey}) vs. Market Share(S_{Turkey})

7.3.2 Stationarity

In our approach we assume that the market share is more or less constant over time, i.e. we have stationary market share series. If the market share data is not stationary, the standard assumptions for asymptotic analysis will not be valid. As a result, the regression parameters cannot be trusted because the t-values do not necessarily follow a t-distribution. Next to this, in a non-stationary series, a shock will have a permanent effect on all future observations. We performed the Augmented Dickey Fuller test to test for a unit root in the market share data. We also included seasonal dummies to test for seasonality. The p-values of the Augmented Dickey Fuller test are presented in table 7.1.

Test	$\mathbf{H_0}$	**Greece**	**Spain**	**Turkey**	**Portugal**
ADF	Unit Root in S_r	0.00	0.00	0.01	0.00

Table 7.1: Unit Root Test Results: p-values of the Augmented Dickey Fuller test

In table 7.1, we see that for all market share series the null hypothesis of a unit root is rejected (with $\alpha = 0.05$). We conclude that none of the market share series contains a unit root.

If a unit root is found one would typically take the first differences of the market shares (ΔS_{rt}) to obtain a stationary series.

7.3.3 Outlier Detection

As a last step in the pre-modeling data analysis we study the dataset to identify possible outliers. As in [22], an outlier can be defined as follows:

> *"An outlier is an observation which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism"*

In the tour operating industry, we know there many mechanisms that influence people to buy certain products. A lot of people are attracted to a product for other reasons than the price. For example, marketing activities attract people to book their holiday in a certain week. From experience we know that marketing activities can have huge effects on the number of bookings.

We determine outliers using the dffits as described in [23] (page 383). Dffits can be seen as "a scale invariant measure for the difference in fitted values". First, we estimate the complete regression model (4.4) including all countries. This model is used to calculate the dffits. We assume the difference in fitted value is significant if the dffits value is larger than $2\sqrt{k/n}$, with k the number of explanatory variables and n the number of observations. We assume that all values that are hard to fit are due to marketing actions or other non-price related shocks in market share. All possible identified outliers are omitted in the final regressions. This mechanism makes sure that the most influential observations do not bias the elasticity estimates. However, one must be careful not to delete observations that contain valuable information.

7.4 Elasticity Estimation

After we analyzed the data extensively in the previous section we can perform the elasticity estimation using the regression model from (4.4). The results are described in this section.

7.4.1 Cross Elasticities

We performed the regression in equation (4.4)[3] with the outlier treatment described in the previous section. We use backward elimination to determine the significant relationships until we obtain the final model in which all parameters are significant (with a α of 0.05). After this, we performed several statistical tests to make sure that the regression assumptions are met. The used tests and assumptions are described section 4.3. The test results of the described tests are presented in table 7.2.

[3] We use GRETL to perform the data analysis

		Models			
Test	$\mathbf{H_0}$	**Greece**	**Spain**	**Turkey**	**Portugal**
ADF	Unit Root in S_r	0.00	0.00	0.01	0.00
RESET	Specification is Adequate	0.63	0.09	0.21	0.93
BP	No Heteroskedasticity	0.76	0.89	0.48	0.72
JB	Normally Distributed Error	0.09	0.10	0.23	0.45
BG	No Autocorrelation	0.00	0.00	0.00	0.79
VIF	multicollinearity G_{Greece}	1.53	1.53	1.53	1.53
VIF	multicollinearity G_{Spain}	1.40	1.40	1.40	1.40
VIF	multicollinearity G_{Turkey}	1.29	1.29	1.29	1.29
VIF	multicollinearity $G_{Portugal}$	1.50	1.50	1.50	1.50

Table 7.2: Null hypothesis and p-values of Statistical Tests

We performed the RESET test to check whether the specified linear models are adequate. Looking at the RESET test results we see that for none of the regions, the null hypothesis of an adequate specification is rejected. Therefore, we can conclude that for all regions the linear specification is adequate. In order to test for heteroskedasticity, we also performed the Breusch Pagan (BP) test. When we look at the results of the BP test, we see that for none of the regions the null hypothesis of no heteroskedasticity is rejected. Therefore, as a result of the BP test we conclude that heteroskedasticity is not a problem. We used the Jarque-Bera (JB) test, to test the normality of the residuals. Looking at the JB test results, we see that for none of the regions the null hypothesis of normally distributed error terms is rejected. Therefore, we can conclude that the error terms are normally distributed. To test for autocorrelation, we performed the Breusch-Godfrey (BG) test. When we look at the results of the BG test, we see that the null hypothesis of no autocorrelation cannot be rejected for the model of Portugal. So for Portugal we can conclude that autocorrelation is not a problem. For all other regions, the null hypothes of no autocorrelation is rejected. For these regions we cannot conclude there is no autocorrelation present. Although serial correlation does not affect the consistency of the estimated regression estimates, it does affect our ability to conduct valid statistical tests. Therefore as discussed in section 4.3, we use Newey-West standard errors which are known to be Heteroskedasticity and Autocorrelation Consistent (HAC). When we look at the Variance Inflation Factors (VIF), we see that all VIF values are smaller than two. A VIF larger than ten is an indication of multicollinearity. Therefore, we conclude that multicollinearity is not a problem in the performed regressions.

We repeat the regression procedure, only now we use Newey-West Standard Errors to account for the found autocorrelation. The final regression results are presented in table 7.3. We also present the standard errors of the regression next to the estimated parameters. In figure 7.5, the residuals are plotted with the corresponding normal distribution fit.

	S_{Greece}	S_{Spain}	S_{Turkey}	$S_{Portugal}$
Constant	0.260 *(0.015)*	0.201 *(0.017)*	0.179 *(0.012)*	0.065 *(0.004)*
Trend	0.000 *(0.000)*	0.000 *(0.000)*	0.000 *(0.000)*	0.000 *(0.000)*
Greece	-0.057 *(0.014)*	-	-0.037 *(0.009)*	-
Spain	-0.056 *(0.016)*	-0.071 *(0.019)*	-0.078 *(0.008)*	-0.016 *(0.004)*
Turkey	-	-	-0.023 *(0.009)*	-0.008 *(0.003)*
Portugal	-	-	-	-0.010 *(0.003)*
# outliers	10	9	10	11
# obs	171	172	171	170
SSE	0.087	0.031	0.017	0.004
R^2	0.549	0.494	0.624	0.519
p-value	0.000	0.000	0.000	0.000

Table 7.3: Regression Results

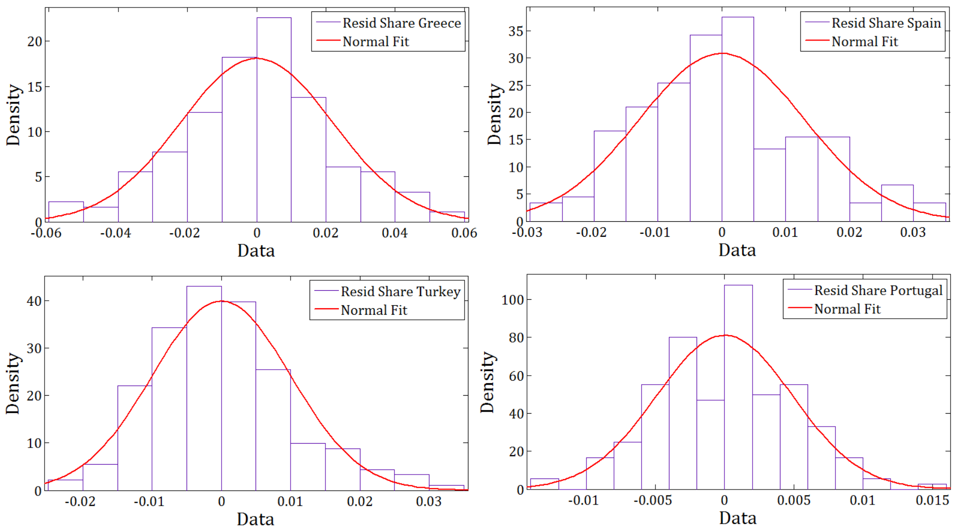

Figure 7.5: Residuals of the regression models including normal pdf fit

To interpret the regression results it is important to understand the equation in (4.4). In (4.4), the Sunweb market share of region r is explained using the price of this region compared to the same region in the market (G_{rt}) and the price of this region compared to other regions within Sunweb (H_{rot}). G_{rt} and H_{rot} are defined in equations (4.2) and (4.3), respectively. The estimated parameters can be interpreted as follows: let $\hat{\beta}_{ro}$ be the parameter estimate, when we increase the relative price of region r compared to region o with one unit, the market share of region r is expected to increase with $\hat{\beta}_{ro}$. For example, when $H_{GreeceSpain}$ is increased with 1 unit, S_{Greece} is expected to increase with -0.056 units.

For the estimated regression models in table 7.3, the constant and trend are significant in all cases. Considering the elasticities, it is remarkable that all elasticities are negative. This means that when the relative price of region r increases compared to the other prices, the expected market share of r decreases. This relationship was already expected, see section 4.1.2. It is also remarkable that the relative price of the region compared to the market price is significant in all models.

Considering the model for Greece, a significant negative elasticity is found for both the price compared to the market as the price compared to Spain. The other relative prices do not have significant influence on the market share of Greece. In the case of Greece 10 outliers are omitted, this leaves 171 remaining observations. An interesting measure for the goodness of fit is the sum of squared error (SSE), in the SSE all regression error terms are squared and summed ($\sum \varepsilon^2$). The SSE is a measure of the discrepancy between the data and the estimated model. In case of Greece, the SSE is equal to 0.087. Another interesting measure is the R^2, this is known as the squared multiple correlation coefficient. R^2 represents the percentage of variability in market share that can be explained by the explanatory variables. The fitted model, has an R^2 of 0.549, this means that 54.9% of the variation in the market share can be explained by this model. We also perform an ANOVA test with the null hypothesis: all coefficients are zero. The alternative hypothesis is: at least one of the coefficients is non-zero. The p-value in the table corresponds to this ANOVA test. For the Greece model, the null hypothesis is rejected. So, we cannot conclude that at least one of the coefficients is non-zero.

In the model for Spain, the only significant elasticity is found for the price compared to the market. The prices compared to other regions do not have significant influence on the market share of Spain. In the case of Spain 9 outliers are omitted, this leaves 172 remaining observations. The SSE is equal to 0.031, this is smaller than the SSE for Greece. This does not necessarily imply a better fit. Since the average share of Greece is larger, the SSE of Greece is expected to be larger as well. The R^2 of the Spain model is equal to 0.494, this means that 49.4% of the variability in market share can be explained by this model. The null hypothesis of the ANOVA is again rejected, so we cannot conclude that at least one of the coefficients is non-zero.

Next, the model for Turkey is considered. Negative elasticities are found for the relative price compared to the market and the relative prices compared to Greece and Spain. The relative price compared to Portugal does not have significant influence on the market share of Turkey. In this model 10 outliers are omitted, the model is based on the 171 remaining observations. The SSE is equal to 0.017, this is smaller than the SSE for Greece and

Spain. The R^2 of the Turkey model is equal to 0.624, meaning that 62.4% of the variability in market share can be explained by the model. The null hypothesis of the ANOVA is again rejected, so we cannot conclude that at least one of the coefficients is non-zero.

Finally, the model for Portugal is considered. Negative elasticities are found for the relative price compared to the market and the relative prices compared to Spain and Turkey. The relative price compared to Greece does not have significant influence on the market share of Portugal. In the Portugal model 11 outliers are omitted, the model is based on the 170 remaining observations. The SSE is equal to 0.004, this is the smallest SSE so far. This is not surprising since the average market share of Portugal is by far the smallest. The R^2 of the model is equal to 0.519, meaning that 51.9% of the variability in market share can be explained by the model. The null hypothesis of the ANOVA is again rejected, so we can again not conclude that at least one of the coefficients is non-zero.

For each of the regression models we also plotted the residuals in figure 7.5, we also included a fit of the normal pdf. According to the log likelihood values of the corresponding fits we can compare the models. The highest log likelihood is found for the Portugal (705.7), after this comes Turkey (577.4), then Spain (530.6) and the lowest log likelihood is found for Greece (434.1). For all models it holds that the JB test does not reject the null hypothesis of normally distributed residuals.

Practical Interpretation

Considering the models in table 7.3, the market share of Spain is the hardest to explain. The R^2 of the Spain model is the lowest of all four models. Next to this, only the relative price compared to the market has a significant influence on Spain's market share. For Greece, only the relative price of Greece compared to the market and Spain has a significant effect on the market share of Greece. For Turkey and Portugal, three relative prices compared to other regions and the market are influencing the market share. It looks like the market shares of the popular regions (Greece and Spain) are harder to influence than the market shares of the less popular regions (Turkey and Portugal). It also looks like the customers to Turkey and Portugal are willing to switch their destination easily if another region or the market is cheaper. Or alternatively, people are willing to go to Turkey and Portugal if the relative prices are favorable. Customers to Greece are only willing to switch to cheaper alternatives in Spain and the market to Greece. The customers to Spain are not willing to switch destinations, however when the market is cheaper they will switch to other trips to Spain in the market.

7.4.2 Elasticities

In this section we already estimated the cross elasticities between the regions. These cross elasticities can be used to solve the first step of the DRDP model. However, in the second step of the DRDP model and the IRDP model the price sensitivity is modeled with the price elasticities of demand. In section 7.3.1 we saw that, due to seasonality, linear least squares cannot be used to estimate the price elasticities of demand from the dataset. Therefore, we create our own dataset based on the cross elasticities from table 7.3. In section 4.2 we explained this process in greater detail. The price elasticities of demand are estimated from this dataset. The distributions, paramater estimates, likelihood values and Kolmogorov Smirnov (KS) p-values are presented in table 7.4.

Panel A: Sunweb Price Distribution Fits

Country	**Greece**	**Spain**	**Turkey**	**Portugal**
Distribution	EV	GEV	GEV	GEV
location: μ	547.0	494.1	512.7	489.6
scale: σ	50.2	42.0	47.6	62.1
shape: ξ	-	-0.185	-0.425	-0.239
Log-Likelihood	-992.7	-947.9	-970.1	-1003.0
KS p-value	0.84	0.76	0.86	0.52

Panel B: Market Price Distribution Fits

Country	**Greece**	**Spain**	**Turkey**	**Portugal**
Distribution	EV	EV	EV	EV
location: μ	707.6	679.4	676.2	656.2
scale: σ	71.2	38.0	63.7	55.8
shape: ξ	-	-	-	-
Log-Likelihood	-1104.8	-968.5	-1058.3	-1015.3
KS p-value	0.11	0.37	0.46	0.29

Panel C: Market Demand Distribution Fits

Country	**Greece**	**Spain**	**Turkey**	**Portugal**
Distribution	GEV	GEV	GEV	GEV
location: μ	1787.5	6075.8	3741.9	614.9
scale: σ	2636.3	1786.5	2039.4	1218.0
shape: ξ	0.033	0.403	0.095	0.203
Log-Likelihood	-1646.6	-1684.6	-1674.7	-1469.4
KS p-value	0.05	0.53	0.07	0.70

Table 7.4: Distribution Fits

We look the distributions and corresponding fits in table 7.4, for all data samples we found a distribution that fits the data properly. The null hypothesis (H_0=The compared samples are drawn from the same distribution) of the KS-test is never rejected. Therefore, we conclude that all fitted distributions fit the data properly. For each fit we also calculated the Log Likelihood of the fitted parameters. Note that we can only compare the results for the same chosen distribution. We see that for the EV distribution, the market price for spain results in the highest likelihood. On the contrary, the market price for Greece results in the lowest log likelihood. For the GEV distributions, the price distribution fits have considerably higher likelihoods than the market demand fits. Generally, the market demand fits are the hardest to fit to the data.

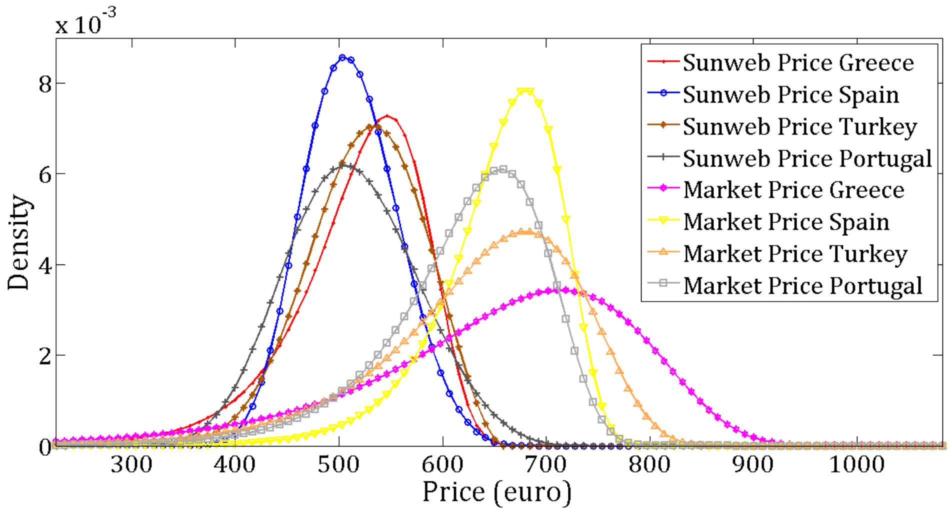

Figure 7.6: Price Distribution Fits

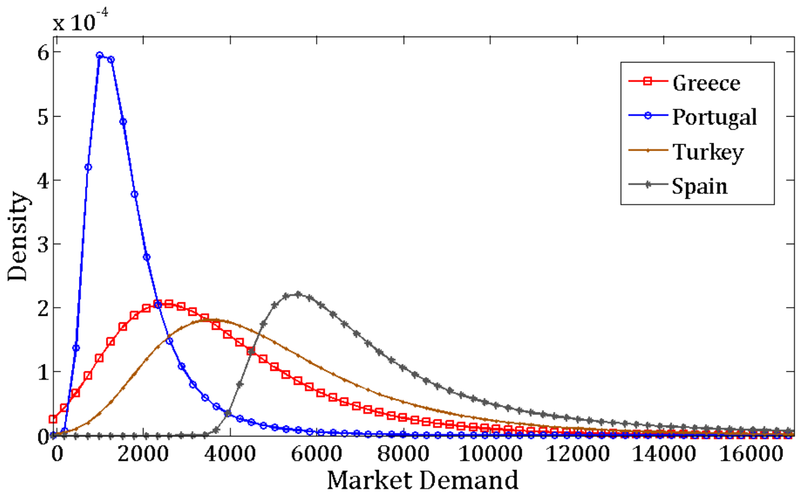

Figure 7.7: Market Demand Distribution Fits

In figures 7.6 and 7.7 we plotted the price and markets demand distributions, respectively. All the market prices from figure 7.6 follow an EV distribution, the Sunweb price for Greece also follows a EV distribution. The Sunweb prices of Spain, Turkey and Portugal from figure 7.6 follow a GEV distribution. The market demand for each of the countries in figure 7.7 follows a GEV distribution. In figure 7.6 we see that the market prices are on average higher than the Sunweb prices.

Remarkable to see that for the sunweb price distribution fits the shape parameter ξ is negative. According to [32], for a negative value of ξ, the GEV distribution belongs to the Weibull Maximum Domain of Attraction and has a finite right endpoint, this finite right endpoint can clearly be seen in figure 7.6. For the market demand distribution fits, the shape parameter ξ is always positive. According to [32], for a positive value of ξ, the GEV distribution belongs to the Fréchet maximum domain of attraction and is regularly varying (power-like tail), this power like tail can clearly be seen in figure 7.7.

We used the fitted distributions from table 7.4 to create the dataset and estimate the price elasticities of demand, as described in section 4.2. The estimated price elasticities of demand are given in table 7.5.

Test	**Greece**	**Spain**	**Turkey**	**Portugal**
Price Elasticity of Demand	-0.067	-0.026	-0.021	-0.051

Table 7.5: Price Elasticity of Demand Estimates

The elasticities of demand presented in table 7.5 are used in the IRDP model and in the second step of the DRDP model. By deriving the price elasticities of demand from the cross elasticities we made sure that the elasticities from both models are generated under the same data generating process without any disturbances. This makes sure that the models can be compared fairly.

7.5 Validation of DRDP Model

In this section we use simulation to compare the results of different pricing models under different circumstances. We define multiple test cases, representing different possible situations. We vary three different attributes of a test case: capacity, margins and demand volatility. The aim is that the test cases provide a realistic mix of different possible situations. First, we compare the models after the first step of the DRDP model. Hereafter, we compare the models after the second step of the DRDP model. To be able to solve the second step of the model, we first have to extend some of the test cases. We are interested in the average profit generated for each of the models under different circumstances. We are also interested in the robustness of the models. Therefore, we also look at the number of simulations that each of the models generates the highest profits. We also look at the number of simulations that each of the models generates the lowest profits. Next to this, we are interested in the average capacity utilization and the number of lost customers.

The simulation outcomes provide us insight in the behavior of the optimization models, they also show the possible value of the DRDP model.

We also discuss the results for the DRDP model using the stochastic constraints described in section 5.4. We are interested in the effect of these constraints on both the profit and the lost demand.

The models in this section are solved, using CPLEX 12.4 as solver[4].

7.5.1 Step 1: Test Cases

In this subsection we describe the basic test cases of the simulations. The test cases are supposed to represent different possible situations. For each test case, we consider 4 traveling regions: Greece, Spain, Turkey and Portugal. The booking horizon consists of one time period with 10000 expected market bookings. We focus on the products departing at the end of this time horizon. The starting prices per region are given table 7.6. From the prices, market prices and market demand we can calculate the expected Sunweb demand for each region ($E(Q_r)$), using formula (5.7).

	Greece	**Spain**	**Turkey**	**Portugal**
Price (€)	519	512	524	512
Market Price (€)	668	657	642	626
$E(Q_r)$	1693	1457	432	323

Table 7.6: Base Situation

We are interested in the behavior of the models under different circumstances. To create varying test cases we focus on three important properties: Capacity, Margins and Demand Volatility. Next, each of these properties with corresponding choices is described.

Capacity/Demand Ratio

We are interested in the behavior of the models under different capacity scenarios. Therefore, we define three capacity situations. The capacity per region can be defined as a ratio of the expected demand in the base situation. The capacity/demand ratio is defined per region in (7.1):

$$Capacity/DemandRatio_r = \frac{TCap_r}{E(Q_r)}, \tag{7.1}$$

[4]Processor: Intel(R) Core(TM) i5-2520M CPU @ 2.50GHz, 2501 Mhz, 2 Core(s), 4 Logical Processor(s)

with $TCap_r$ the capacity for products to region r. And $E(Q_r)$ the expected demand in the base pricing decision, given in table 7.6. We define three different capacity situations with capacity/demand ratios of 1.5, 1.2 and 0.9 for high, average and low capacity situations respectively.

Margins

We are interested in the behavior of the models under different margin scenarios. We expect that the DRDP model might be able to exploit the fact that the margins of some regions are higher than the margins of other regions. Therefore, we set different costs per region to be able to control the margins of the products. We define an equal margins situation in which the base margins of all regions are equal to €100. We define an unequal margin situation in which the base margins of the regions differ. In this situation we increase the base margins of Greece and Turkey to €150 and we decrease the base margins of Spain and Portugal to €50.

Demand Uncertainty

We are interested in the robustness of the different models. Therefore, we also define situations with different demand volatility. We refer to section 5.4 for a discussion on demand uncertainty in RM models.

After the models are solved, we simulate the actual demand using the expected demand as mean and a chosen standard deviation. We define 2 possible choices for the standard deviation, a high standard deviation or a low standard deviation. The high and low standard deviations are determined as a function of the expected demand per region. In the high demand volatility case, the demand volatility is defined as $\sigma_r = 0.25E(Q_r)$. In the low demand volatility case, the demand volatility is defined as $\sigma_r = 0.1E(Q_r)$. We assume that for each of these situations the demand volatility is known beforehand.

Summarizing Test Cases

In table 7.7 we provide an overview of all the Test Cases. We look at all combinations of the described test case properties. The computational results are described next, in section 7.5.2.

7.5.2 Step 1: Computational Results

We solved step 1 of the DRDP model for each of the test cases described in table 7.7. The optimization starts with the prices of the products before optimization, these are called the base prices. In the results, the Base pricing decision is also considered. For each of the test cases, we also solved

	Capacity	**Margins**	**Demand Uncertainty**
Test Case 1	High	Equal	High
Test Case 2	High	Equal	Low
Test Case 3	High	Unequal	High
Test Case 4	High	Unequal	Low
Test Case 5	Average	Equal	High
Test Case 6	Average	Equal	Low
Test Case 7	Average	Unequal	High
Test Case 8	Average	Unequal	Low
Test Case 9	Low	Equal	High
Test Case 10	Low	Equal	Low
Test Case 11	Low	Unequal	High
Test Case 12	Low	Unequal	Low

Table 7.7: Different Test Cases Step 1

the IRDP model. After the models are solved, we have the pricing decisions for each of the models with corresponding expected demands. The realized demand for each region is simulated as random number from the normal distribution. We use the expected demand as mean and 0.1 and 0.25 times the expected demand as standard deviations for the certain and uncertain demand situations respectively. For each test case, 10000 simulations are performed. The complete computational results are shown in table 8.1 (test cases 1-3), table 8.2 (test cases 4-8) and table 8.3 (test cases 9-12) in Appendix I.

Average Results

In table 7.8, the average results over all 12 test cases are presented. These results represent the average performance of the models over 12 different test cases. For each test case we performed 10000 simulations, so in total these results are based on 120000 simulations.

	Base	**IRDP**	**DRDP**
Avg Profit (€)	358934	438355	523077
Std Profit (€)	36498	41448	55441
Profit as % of IRDP	82.0 %	100.0 %	119.9 %
Profit % Highest	0.0 %	1.7 %	98.3 %
Profit % Lowest	88.5 %	10.8 %	0.7 %
Cap % Utilization	81.6 %	80.8 %	79.2 %
Avg Demand Lost	190.3	180.1	114.4

Table 7.8: Average Results of all 12 Test Cases

The most important benchmark is the average realized profit. The base pricing decision leads to the lowest average profit of €358934, with a standard deviation of €36498. In 0.0% of the simulations, the base pricing decision led to the highest realized profit. In 88.5% of the simulations, the base pricing decision led to the lowest realized profit. Using the base prices led to an average capacity utilization of 81.6% and an average lost demand of 190.3 customers.

The IRDP pricing decision leads to an average profit of €438355, with a standard deviation of €41448. In 1.7% of the simulations, the IRDP pricing decision led to the highest realized profit. In 10.8% of the simulations, the IRDP pricing decision led to the lowest realized profit. Using the IRDP prices, led to an average capacity utilization of 80.8% and an average lost demand of 180.1 customers.

Finally, the DRDP pricing decision leads to the highest average profit of €523077, with a standard deviation of €55441. In 98.3% of the simulations, the DRDP pricing decision led to the highest realized profit. In 0.7% of the simulations, the DRDP pricing decision led to the lowest realized profit. Using the DRDP prices, led to an average capacity utilization of 79.2% and an average lost demand of 114.4 customers.

Note that the standard deviation of the profits is the highest for the DRDP model. As expected, the standard deviations of the profits are much smaller in case of low demand uncertainty compared to high demand uncertainty.

For each model, the profit is expressed as a percentage of the IRDP profit. Over all test cases, the DRDP model realized on average 19.9% more profit than the IRDP model. Not only the average profits are higher, the percentage of the simulations the DRDP model performs best is also quite convincing. Of all three pricing decisions, the DRDP model realized the highest profit in 98.3% of the simulations while the IRDP model realized the highest profit in only 1.7% of the simulations. The DRDP model realized the lowest profit in 0.7 % of the simulations, for the IRDP model this is 10.8% of the simulations. From these percentages we can conclude that the DRDP model is also quite robust.

Compared to the IRDP model, the DRDP model realized a lower capacity utilization and less lost demand. This implicates that the average prices in case of the DRDP model are higher than for the IRDP model.

Next, we discuss the results for the different capacity, margins and demand volatility situations described in section 7.5.2.

Low, Average and High Capacity

Three different capacity situations are considered. In the first situation, the capacity is considered high (1.5 times the expected base demand). In the second situation, the capacity is considered average (1.2 times the expected base demand). In the third situation, the capacity is considered low (0.9

times the expected base demand). For each of these situations, the realized profits for all test cases were aggregated. The average realized profits for each of the models, for the different capacity situations, are presented in figure 7.8.

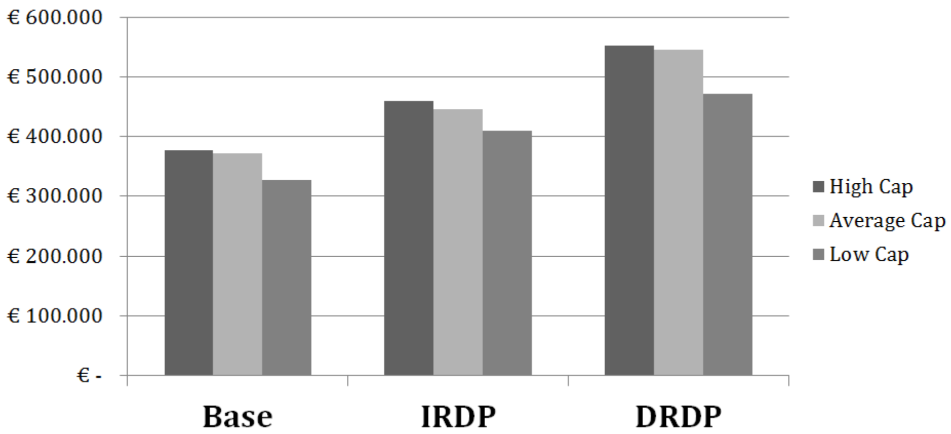

Figure 7.8: Average Profit for Low, Average and High Capacity

For all models, it holds that a higher capacity ratio leads to a higher average profit. With higher capacities, more products can be sold in the optimization period. This leads to higher average profits and less lost demand. The Profit difference between the high and average capacity situations is smaller than between the average and low capacity situations. In the lowest capacity situations, the base capacity is smaller than the expected demand. In that case, a unit of additional capacity is more valuable than in the case of average capacity.

Comparing the optimization models, the DRDP model realized a higher average profit than the IRDP model for each of the different capacity/demand situations. Next to this, the DRDP realized the highest profit in 98.7%, 99.1% and 97.4% of the simulations for the high, average and low capacity ratios, respectively. The DRDP realized the lowest profit in 0.4%, 0.4% and 1.7% of the simulations for the high, average and low capacity ratios, respectively. We conclude that the pricing of the DRDP model is preferred in all capacity situations.

Equal and Unequal Margins

We considered different margin scenarios. In the first scenario, all regions have an equal base margin of €100 per product. In the second scenario, we increase the base margins of Greece and Turkey to €150 and we decrease the base margins of Spain and Portugal to €50. We expected that the DRDP model might be able to exploit the fact that the margins of some regions are higher than the margins of other regions.

We aggregated all test cases with equal base margins and compared them to the test cases with unequal base margins. The realized profits for both margin situations, for all models are shown in figure 7.9.

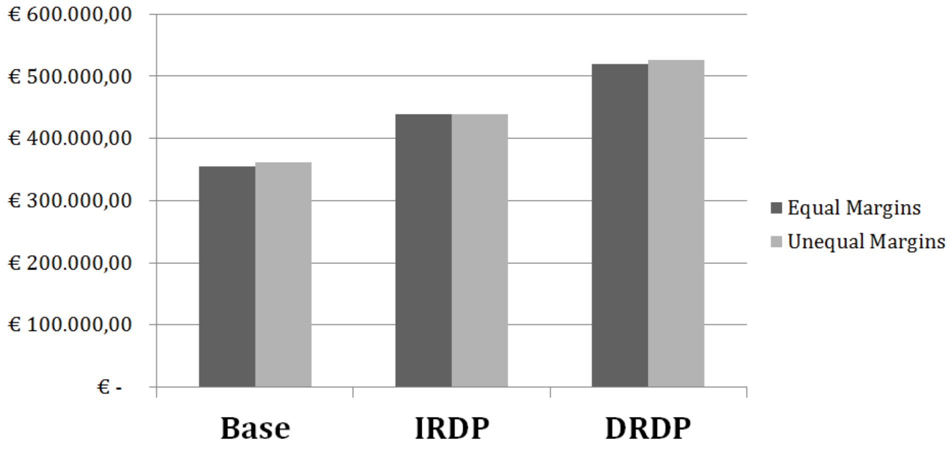

Figure 7.9: Average Profit for Equal and Unequal Margins

For both the base pricing decision and the DRDP pricing decision, the average profit is 1.0% higher in case of unequal margins. Nevertheless, the differences are minimal. For the IRDP model, the average realized profit is more or less equal for both situations (0.0% difference). No large differences are found between the equal margin situations and the unequal margin situations.

Comparing the profits generated under the different optimization models, the average profit under the DRDP pricing policy is higher for both equal and unequal margins. The DRDP model realized the highest profit in 98.1% and 98.7% of the simulations for the equal and unequal margins, respectively. The DRDP model realized the lowest profit in 0.5% and 0.8% of the simulations for the equal and unequal margins, respectively. For both equal and unequal margin situations, the DRDP pricing policy is preferred.

Low and High Demand Volatility

We considered the situation of high demand volatility (0.25 times the expected demand) and low demand volatility (0.1 times the expected demand). We aggregated all test cases with high demand volatility and compared them to the test cases with low demand volatility. The realized profit for both volatility situations, for all models is shown in figure 7.10.

All models realize higher average profits under low demand volatility compared to high demand volatility. When the demand becomes more volatile, the demand estimates used in the models become less accurate.

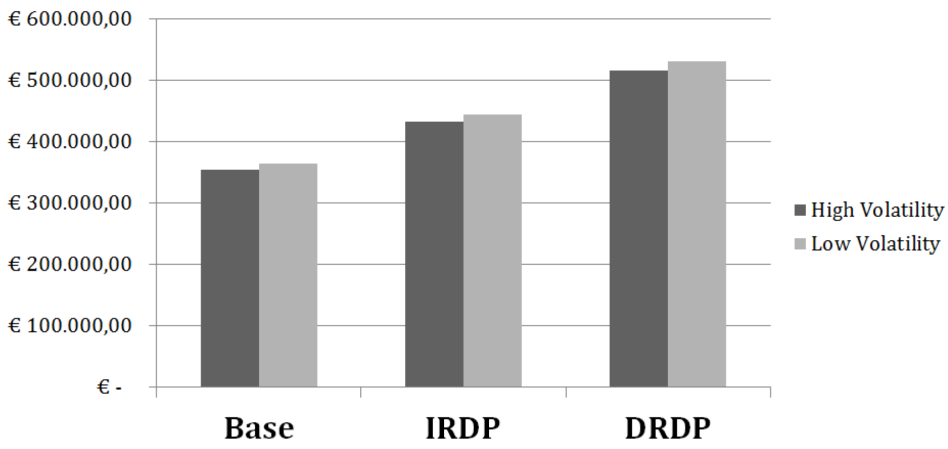

Figure 7.10: Average Profit for High and Low Volatility

Large underestimations of demand lead to large numbers of customers lost, large overestimations of demand lead to a large number of unsold capacities. Both these situations result in lost potential profit. Looking at the average realized profit, none of the models in specific seem to benefit from higher or lower demand volatility.

The DRDP model generated a higher average profit for both low and high demand volatility situations. In case of low demand volatility the DRDP model realized the highest profit in 99.8% of the simulations. In case of high demand volatility, the DRDP model realized the highest profit in 96.9% of the simulations. The DRDP model realized the lowest profit in 1.3% and 0.0% of the simulations, for high and low demand volatility, respectively. The DRDP model benefits from the more accurate information, in case of low demand volatility the DRDP model realized the highest profit in almost all simulations. In both low and high demand volatility situations, the DRDP pricing policy is preferred.

Interaction Effects

So far we focused on the individual effects of the capacity, margings and demand volatility situations. However, interaction effects may also be present. To study these, we take a closer look at the average realized profits from the complete computational results shown in table 8.1 (test cases 1-3), table 8.2 (test cases 4-8) and table 8.3 (test cases 9-12) in Appendix I. In figure 7.11 we plotted the average realized profits for each test case individually.

We see that in case of high capacity, the interaction effects are minimal. For each test case with high capacity (test cases 1-4), the realized profits are comparable. Only in case of unequal margins, the IRDP realized slightly less profit and the DRDP model realized slightly more profit. In case of

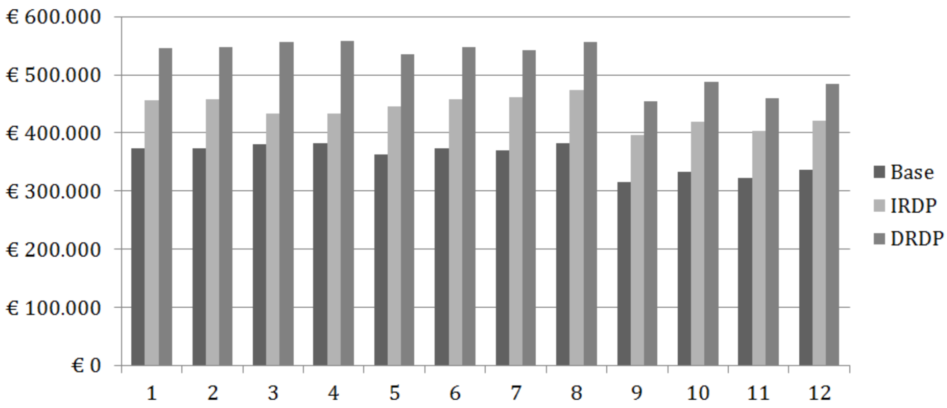

Figure 7.11: Realized average profits for all test cases

average capacity (test cases 5-8), we see that the combination of low demand volatility and unequal margins leads to the highest profits for both the IRDP and DRDP model. In case of low capacity (test cases 9-12), the demand volatility becomes more important. The gaps between the realized profits of high demand volatility and low demand volatility situations is the largest in this case. In case of low capacity, extra unforseen demand might lead to large lost profits. Whereas, in case of high capacity the extra unforseen demand still fits within the capacity.

7.5.3 Step 2: Test Cases

In section 7.5.1, we described the regional test cases used to test step 1 of the DRDP model. Because the data of all regional clusters is aggregated, both models are solved as if there is only one product per region. This is not completely fair because the DRDP model is designed for this purpose. Whereas, the IRDP model is designed for the pricing of multiple products that make use of the same resources. Next to this, DRDP model solves this pricing problem of the regional clusters to optimality. To be able to make a fair comparison between the models, we are also interested in the results after step 2 of the DRDP model. For solving step 2 of the DRDP model, we need to a situation with multiple products per region. To create this situation, we expand some of the previously described test cases.

We focus on the optimization of products to Greece. To create different products we define different accommodation types and different possible departure dates. The total demand for trips to Greece is split into demand for these products within Greece. To look at different problem sizes we also define two scenarios. In the first scenario we have ten possible departure dates and in the second scenario we have hundred possible departure dates. In both scenarios there are three different accommodation types. This results in 30 different products for the small problem 300 different products for the large problem.

Departure dates

We define multiple departure dates from which the arriving customers can choose. Note that the base demand for Greece is already known from table 7.6 (1693 customers). We draw random numbers from a multinomial distribution to distribute the region demand over the different departure dates. The Greece base demand is used as the number of trials and the probability of success for each departure date is equal. All trips are 8 days and every 3 days a flight departs.

Accommodations

We define three types of accommodations: cheap ones, average ones and expensive ones. An 8 day trip to an average priced accommodation costs €519, this is equal to the Greece base price from table 7.6. A trip to the cheap accommodation is 20% cheaper and a trip to the expensive accommodation is 20% more expensive than base price. The demand for the different types of accommodations is equally distributed, each accommodation type gets one third of the total demand. For each of the accommodations, inbound flights and outbound flights the capacity-demand ratio is 1.2.

Test cases 5 and 6 from table 7.7 are expanded. This means we are still interested in the different demand volatility scenarios. An overview of the expanded test cases is given in table 7.9.

	Demand Volatility	**# departure Dates**	**# Acco's**
Test Case 5.1	High	10	3
Test Case 5.2	High	100	3
Test Case 6.1	Low	10	3
Test Case 6.2	Low	100	3

Table 7.9: Expanded test cases 5 and 6

7.5.4 Step 2: Computational Results

In the previous section, we described the expanded test cases. For each of these test cases, the IRDP model and the DRDP model are solved. Again, we consider the Base pricing decision, this is the pricing decision before optimizing the pricing policy. The base price of an average priced accommodation is €519. A trip to a cheap accommodation is 20% cheaper and trip to an expensive accommodations are 20% more expensive. After the models are solved, we again simulate the actual realized demand. For each product, we use a normal distribution with the expected demand as mean and 0.1 and 0.25 times the expected demand as standard deviations for the low and high demand volatility situations, respectively.

The complete computational results are presented in table 8.4 in Appendix II.

Average Results

We aggregated the computational results from the different test cases. The average results are presented in table 7.10. For each test case we used 10000 simulations, so these results are based on a total of 40000 simulations. Note that these are the results for Greece only.

	Base	**IRDP**	**DRDP**
Avg Profit (€)	191599	261923	280564
Std Profit (€)	5477	8312	8944
Profit % of Base	73.2 %	100.0 %	107.1 %
Profit % Highest	0.0 %	12.1 %	87.9 %
Profit % Lowest	85.7 %	10.3 %	4.1 %
Cap % Utilization	90.1 %	89.4 %	88.8 %
Avg Demand Lost	80.2	76.1	69.0

Table 7.10: Average Results of Expanded Test Cases

The most important benchmark is the average realized profit. The base pricing decision leads to the lowest average profit of €191599, with a standard deviation of €5477. In 0.0% of the simulations, the base pricing decision led to the highest realized profit. In 85.7% of the simulations, the base pricing decision led to the lowest realized profit. Using the base prices led to an average capacity utilization of 90.1% and an average lost demand of 80.2 customers.

The IRDP pricing decision leads to an average profit of €261923, with a standard deviation of €8312. In 12.1% of the simulations, the IRDP pricing decision led to the highest realized profit. In 10.3% of the simulations, the IRDP pricing decision led to the lowest realized profit. Using the IRDP prices led to an average capacity utilization of 89.4% and an average lost demand of 76.1 customers.

Finally, the DRDP pricing decision leads to the highest average profit of €280564, with a standard deviation of €8944. This is 7.1% higher than the average profit realized after using the IRDP model. The standard deviation is slightly higher than for the IRDP model. In 87.9% of the simulations, the DRDP pricing decision led to the highest realized profit. In 4.1% of the simulations, the DRDP pricing decision led to the lowest realized profit. Using the DRDP prices led to an average capacity utilization of 88.8% and an average lost demand of 69.0 customers. Compared to the IRDP model, the DRDP model again leads to lower capacity utilization. This implicates a higher average price for the products under the DRDP model. On average, these slightly higher prices also result in less lost demand. On average over all expended test cases, the DRDP model is preferred.

Compared to the region results from section 7.5.2, the DRDP pricing decision performed equally well compared to the base pricing decision. In both cases, the DRDP pricing decision leads to an increase of 46% in average profit. The IRDP model performed relatively better in case of multiple products per region. In case of multiple products per region, the IRDP pricing decision realized a 37% higher profit than the base pricing decision. In the case of the regional clusters in table 7.8, this was 22.0%. With multiple products per region, the IRDP model can exploit the fact that products using the same capacities have different demands and prices. Therefore, the IRDP prices result in higher profits compared to the base prices.

Finally, we conclude there are no large differences between the small and large test cases. Next, we discuss the results for the different demand volatility situations.

Low and High Demand Volatility

We considered the situation of high demand volatility (0.25 times the expected demand) and low demand volatility (0.1 times the expected demand). We aggregated all expanded test cases with high demand volatility and compared them to the expanded test cases with low demand volatility. The realized profit for both volatility situations, for all models is shown in figure 7.12.

Like for the regional test cases in section 7.5.2, all models realize higher average profits under low demand volatility compared to high demand volatility. When the demand for the products becomes more volatile, the demand estimates used in the models become less accurate. Large underestimations of demand lead to a higher number of lost customers, large overestimations of demand lead to a higher number of unsold capacities. Both these situations result in lost potential profit. Therefore, the test cases with high demand uncertainty have a higher standard deviation of the profit. Again, none of the models in specific seems to benefit more from higher or lower demand volatility. The DRDP model generated a higher average profit for both low and high demand volatility situations. Considering the highest realized profits per simulation: the DRDP model realized the highest profit in 85.5% and 90.3% of the simulations for high and low demand volatility, respectively. Whereas, the IRDP model realized the highest profit in 14.5% and 7.9% of the simulations for high and low demand volatility, respectively. Considering the lowest realized profits per simulation: the DRDP model realized the lowest profit in 8.2% and 0.0% of the simulations for high and

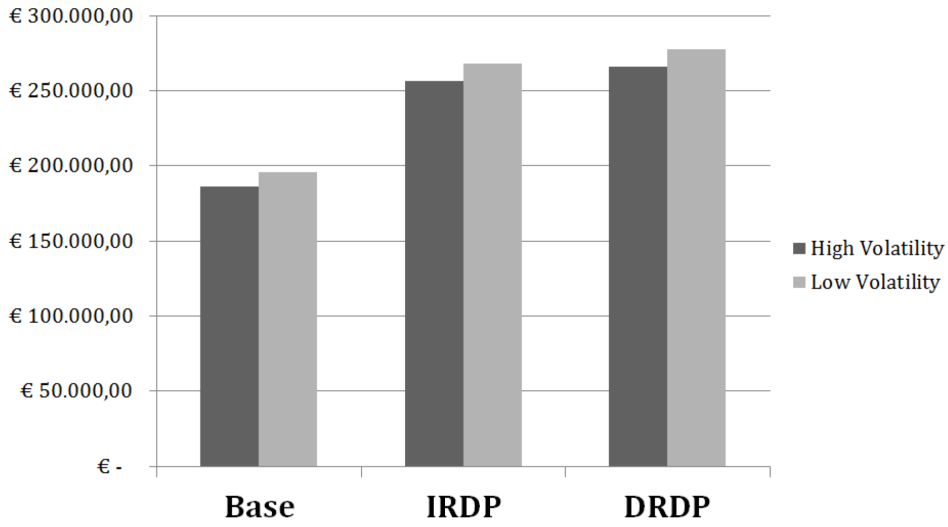

Figure 7.12: Avg Profit for Different Demand Volatility Situations

low demand volatility, respectively. Whereas, the IRDP model realized the lowest profit in 20.5% and 0.0% of the simulations, for high and low demand volatility, respectively.

Again, the DRDP model benefits from the more accurate information. In both low and high demand volatility situations, the DRDP pricing policy is preferred.

7.5.5 IRDP vs DRDP model

In order to compare the IRDP and the DRDP models, we performed a t-test to test the realized profits. The central limit theorem ensures that the (parametric) t-test test works well with large samples, in our case 10000 simulations is sufficient to trust the t-test outcomes. The t-test, tests the null hypothesis that the realized profits are independent random samples from normal distributions with equal means and equal but unknown variances, against the alternative that the means are not equal. For all test cases and expanded test cases, the null hypothesis is rejected (for α=0.01). In case the null hypothesis is rejected we conclude that the model with the highest average profits realized significant higher profits. Therefore, we can conclude that the DRDP model realized significant higher profits than the IRDP model for all test cases.

7.5.6 Stochastic Constraints

In section 5.4 we introduced constraints that can be used to deal with stochastic demand. Instead of existing techniques from amongst [7] and

[51], we proposed a different approach. We use techniques from inventory management to manage the available capacities. Using the proposed constraints we are able to regulate the probability of lost demand. By means of simulation experiments we test whether this approach provides the desired results.

Step 1 of the DRDP Model

In the step 1 of the DRDP model, the stochastic constraint regulates the expected lost demand on the region level. For each of the instances in table 7.7, we solved the DRDP model for $\theta_r = 0.3$, $\theta_r = 0.5$ and $\theta_r = 0.7$. The results are presented in table 8.5 (test cases 1-6) and 8.6 (test cases 7-12) in appendix III.

For the high capacity situations (test cases 1-4), we see that the realized profits are more or less the same for all values of θ_r. Next to this, each model realized the highest profit in approximately 33.3% of the simulations. This also holds for the lowest realized profits. This can easily be explained, by the fact that the optimal pricing decision is feasible for all considered values of θ_r. In fact, the optimzal pricing decision for the different values of θ_r is exactly the same. This means that the required safety buffer for the different values of θ_r does not influence the optimal pricing policy. Moreover, in case of an increasing value of θ_r, the safety buffer declines (in case of $\theta_r = 0.7$ it is even negative). A result of this, is that the capacity constraint is loosened. Therefore, the expected profit is never lower than for a lower value of θ_r.

For the average capacity situations (test cases 5-8), we see that the models with $\theta_r = 0.5$ and $\theta_r = 0.7$ again perform equally well. In fact, the pricing decision again is exactly the same for these models. When we use the value of $\theta_r = 0.3$, a more strict capacity constraint is constructed. In this case, all region prices are set to a maximum (+10%). This is required to fulfill the lost demand probability. We see that this has the expected results for capacity utilization and the lost demand. In all cases, due to the capacity buffer, the realized capacity utilization decreased. This capacity buffer also reduces the lost demand significantly. Looking at the lost demand, we can conclude that the demand buffer brings the desired effect.

On the other hand, on average the higher prices also result in lower profits. Whereas, the DRDP model (with $\theta_r = 0.5$) optimized the prices using the expected demand, the DRDP model with $\theta_r = 0.3$ uses the expected demand including the buffer. On average, over 10000 simulations, the expected demand will be a better estimate for the realized demand than the expected demand plus buffer. So, on average the $\theta_r = 0.3$ model is not preferred. However, there are also cases in which the realized demand turns out to be higher than the expected demand. In these cases, the model with $\theta_r = 0.3$ can still serve all customers for even higher prices per customer. This can be seen in the percentages that this model realized the highest profits. This is approximately 20% and 13% in case of high and low demand volatility, respectively. On the contrary, when the demand does not turn out to be higher than expected the model with $\theta_r = 0.3$ leads to poor results with even lower demands and also lower profits. We can conclude that in the test cases of average capacity, the stochastic constraints are not preferred in terms of profit.

For the low capacity situations (test cases 9-12), we again see that the models with $\theta_r = 0.5$ and $\theta_r = 0.7$ perform equally well. Remarkably, there are no results for the model with $\theta_r = 0.3$. This is due to the fact that the model is not feasible in these cases. Since we use the constraint that the complete price path has to be defined (5.16), the service level from constraint (5.32) cannot be guaranteed. In practice it is very simple to overcome this problem. The equality in constraint (5.16) makes sure that the price path is completely defined by setting it equal to one. When we replace this equality sign by a smaller or equal to sign, the price path does not necessarily have to be completely defined. In order to realize maximum profits the price path will always be completely defined, unless the capacity constraint is violated otherwise. When the price path cannot be completely defined, it is obvious that the considered product has to have the highest price possible (+10%). In case a certain product runs out of capacity it is simply removed from the possible products. For tour operators, this happens all the time especially when the products are close to departure. Note that this practical feature of the price path constraint is already used in the IRDP model.

Step 2 of the DRDP Model

In the step 2 of the DRDP model, the stochastic constraint regulates the expected lost demand on the resource level. For each of the instances in table 7.9, we solved the DRDP model for $\zeta_k = 0.3$, $\zeta_k = 0.5$ and $\zeta_k = 0.7$. The results are presented in table 8.7 in appendix VI.

In general, the results for the stochastic constraints are comparable for the first and second step of the DRDP model (looking at test cases 5-8, with average capacity). For all expanded test cases (5.1-6.2) it holds that the models for $\zeta_k = 0.5$ and $\zeta_k = 0.7$ perform equally well. The pricing policy that is optimal for the $\zeta_k = 0.5$ model can also be obtained by the $\zeta_k = 0.7$ model. Again, the $\zeta_k = 0.3$ model leads to a lower capacity utilization and lower lost demand due to the capacity buffer. Again, the pricing decision under the smaller probability of lost demand leads to significant lower profits. Therefore, to gain maximum profits the original constraints are preferred.

Summarizing, we conclude that introducing buffer capacity, a common used inventory management technique, does reduce lost demand. We can obtain and maintain a desired service level by using a capacity buffer. However, these techniques only twist the prices to obtain a certain service level. In terms of profits, they are not beneficial. We saw those results for stochastic constraints on both the region level and the resource level.

7.6 Computation Times

In sections 7.5.2 and 7.5.4 we solved multiple instances of step 1 and step 2 of the DRDP model, respectively. For both step 1 and step 2 of the model, we discuss the corresponding computation times[5].

Step 1 of the DRDP Model

For the step 1 test cases, we solved the first step of the DRDP model. These basic test cases did not lead to high computation times. All test cases were solved within 0.01 seconds for the DRDP model and 0.001 seconds for the IRDP model. We considered 4 regions, and 5 alternative prices per region. Because our problem instances are very small we cannot conclude that this is the case for larger problem instances. However, the DRDP model is designed in such a way that the first step of the model is never expected to grow very fast. This is because the size of the first step of the model only grows in the number of regions, the number of price alternatives and the number of time periods, this is described in section 5.5.1. The number of time periods and the number of alternative prices can be controlled by the user. In practice, one time period and 5 alternative prices are used. The number of regions is not expected to grow very large.

Step 2 of the DRDP Model

For the step 2 test cases, we solved the second step of the DRDP model. This took 0.0012 seconds for the small problem instance (30 products), for the IRDP model this was 0.0010 seconds. For the large problem instances (300 products) we observed a computation time of 0.0055 seconds. For the IRDP model this was 0.0034 seconds. The computation times of the second step of the DRDP are comparable to the IRDP model. The only difference is the average price constraint, this leads to slightly higher computation times. We expect that larger problem instances can be solved within reasonable time. Moreover, the IRDP model currently solves the actual problem instances (20+ million products) within several minutes.

[5]Note, in AIMMS the computation times are rounded to two decimal places.

Chapter 8

Conclusions and Further Research

8.1 Conslusions

Due to the rising use of internet, the business of e-commerce has grown rapidly in the last years. The easy access to information leads to transparent markets with high competition due to high price sensitive customers. Not only the customers benefit from easy access to information, companies can also benefit from the available market information which is available in many sectors.

In this research we proposed the Dependent Region Dynamic Pricing (DRDP) model. The DRDP model can be used for the dynamic pricing decision of products which make use of both flight and hotel capacities. The DRDP model integrates both market information and competition between regions into the tour operator's pricing decision. The DRDP model is an extension of the Independent Region Dynamic Pricing (IRDP) model, proposed in [50].

In Chapter 1 we provide a short introduction and provide some background information. Hereafter we state the research questions and sub questions.

In Chapter 2, some insights in the actual problem are provided. We provide an overview of the decision making process for the online setting of the pricing problem. This overview enables us to place the RM model into the bigger picture of the tour operator's decision making process. One of the most important things is the composition of the travel products the tour operator offers. We see that these products are competing for space among multiple flights and hotels, this results in a considerable combinatorial problem. [50] tackled this problem using the Independent Region Dynamic Pricing (IRDP) model. We formulate and extensively study the IRDP model. Usually products to different regions do not use overlapping capacities. Next to this, [50] assumed that the pricing decisions of the regions do not influence the demand of other regions. Due to this assumption, the problem can be decomposed into independent sub-problems per region. Decomposing the problem into independent sub-problems considerably reduces the complexity of the problem. Next, we focus on the way consumer behavior is modeled in the IRDP model. In the IRDP model, the demand for a product only depends on the price of this product. We identify that in reality, demand shocks can be caused by price changes of the product, price changes of substitutable products, shocks in market demand and other external factors.

In Chapter 3, the relevant literature is studied extensively. This literature study helps us placing our research into perspective. Not much of this literature proposes RM models specifically for tour operators. Besides this, pricing models from the airline and hotel industries are mostly not directly applicable for a tour operator. The combinatorial effects resulting from products competing for limited hotel and airline capacities result in a completely different problem structure. Therefore, as stated in [50], RM models from the airline and hotel industry do not fit the tour operator business properly. Next to this, several researchers mention the necessity of taking into account the cross price elasticities (for example, [14] and [47]). Therefore, in Chapter 4 and 5 we propose methods to integrate extended consumer behavior into a tour operator's dynamic pricing model.

In Chapter 4, the observed consumer behavior is measured and quantified. The consumer behavior is captured in a multiple regression model. This multiple regression model includes cross regional effects and the external market information. We use the market shares and relative prices to model the characteristics of a transparent market in which customers can easily compare prices. The goal of the regression model is to estimate the expected market share for a certain region, after a given pricing decision. In order to test whether the parameter estimates are valid, in section 4.3 seven regression assumptions with corresponding statistical tests are discussed. For each assumption, there is a different impact on the parameter estimates in case the assumption is not met.

In Chapter 5, the quantified consumer behavior is integrated into a Dependent Region Dynamic Pricing (DRDP) model. The DRDP model contains two stages, in the first stage the cross effects between regional clusters are used to identify the optimal region price. This price is taken into account while pricing the individual products in the second stage. The first step can be seen as the missing link between the individual product optimization and the dependencies between the different regional clusters. The first step is modeled as a Binary Quadratic Programming (BQP) problem, which are usually NP-hard and hence practically difficult to solve. Therefore, we also describe how the quadratic objective can be linearized. The optimal average price per region from step 1 is fixed during the second step of the optimization. Therefore, the only dependency between the regions is fixed. This means that we can again assume independent regions in the second step of the DRDP model. In addition to the IRDP model, we use a constraint to fix the optimal region prices from the first step of the DRDP model. Despite this, the second step of the DRDP model is very similar to the IRDP model. Finally, we propose additional constraints that can be used to deal with stochastic demand. Instead of existing techniques, we proposed a different approach. We use techniques from inventory management to manage the available capacities. Using the proposed constraints we are able to regulate the probability of lost demand.

In Chapter 6, we provide an analytical comparison between the IRDP model and the DRDP model. First, we focus on the number of price combinations of the different models. Among these price combinations, the model has to choose the most profitable pricing decision. The number of price combinations is used as a benchmark for the complexity of the models. We conclude that the difficulty of the problem mainly depends on the number of regions, the number of products per region and the number of alternative prices per product. The size of the first step of the DRDP model only depends on the number of regions and the number of alternative prices per region. Since these two sets are not expected to grow very large, the first step of the DRDP model will also not grow very large. The size of the second step of the DRDP model grows as fast as the size of the IRDP model. The size of the problem in case of the DRDP model will always be larger than the size of the problem in case of the IRDP model. However, the two-step DRDP approach effectively includes cross region effects with minimal increase in complexity. Next, we focus on the comparison of the objectives of the different models. We conclude that in both the objective of the IRDP model and the objective of the second step of the DRDP model, the cross regional profit is missing. We also conclude that, due to the fixed cross regional effects, this did not affect the actual pricing decisions. However, in order to obtain the correct expected profit an adjustment to the objectives is required.

In Chapter 7, we provide a case study on Sunweb, a large online tour operator based in the Netherlands. In a pre-modeling data analysis, we concluded that the price demand relationship is disturbed by high seasonality. In the beginning of the booking season, the prices and demand are relatively high. While, in the end of the booking season, the prices and demand are relatively low. Both these situations cause a positive relationship between price and demand. We conclude that our approach of using the relative prices to explain the market shares is much more stable. Next to this, we found that all of the market share series are stationary over time. In the travel industry, people are influenced by many factors other than price. Marketing activities can cause large demand peaks that cannot be explained by price changes. To be able to determine the effects of price changes on demand, outlier detection is very important. In the elasticity estimation, we found around ten outliers for each of the regions. The remaining observations were used to estimate the relationship between relative prices and market share. We found that the market share of Spain is the hardest to explain, only the relative price compared to the market has a significant influence on Spain's market share. For Greece, only the relative price of Greece compared to the market and Spain has a significant effect on the market share of Greece. For Turkey and Portugal, three relative prices compared to other regions and the market are influencing the market share. It looks like the market shares of the popular regions (Greece and Spain) are harder to influence than the market shares of the less popular regions (Turkey and Portugal). It also looks like the customers to Turkey and Portugal are willing to switch their destination easily if another region or the market is cheaper. Or alternatively, people are willing to go to Turkey and Portugal if the relative prices are favorable. Customers to Greece are only willing to switch to cheaper alternatives in Spain and the market to Greece. The customers to Spain are not willing to switch destinations, however when the market is cheaper they will switch to other trips to Spain in the market.

Next, we used simulation to compare the results for different pricing policies. We compare the base pricing decision, with the optimized pricing decisions from the IRDP and DRDP models. We created multiple test cases to test both the first and the second step of the DRDP model compared to the IRDP model. We varied the capacity, margins and demand uncertainty in order to obtain a realistic mix of different possible test cases. For each test case, we solved both the IRDP and DRDP model. We first focus on the regional pricing decision after the first step of the model, the DRDP model realized on average 19.9% more profit than the IRDP model. The DRDP model led to the highest realized profit in 98.3% of the simulations and to the lowest profit in 0.7% of the simulations. Next to this, for each of the capacity/demand, margin and demand uncertainty situations did the DRDP model lead to the highest average profit. This shows that the DRDP model is highly preferred for the regional pricing decision.

The first step of the DRDP model is designed for the pricing decision of regions. Whereas, the IRDP model is designed for the pricing of multiple products that make use of the same resources. Therefore, in order to make a fair comparison we also compared the IRDP and DRDP model for problem instances with multiple products per region. We found that the DRDP on average led to 7.1% more profit than the IRDP model. Next to this, the DRDP model led to the highest realized profit in 87,9% of the simulations and to the lowest profit in 4.1% of the simulations. Next to this, for each of the demand uncertainty situations did the DRDP model lead to the highest average profit. This shows that the DRDP model is also preferred for the product pricing decision.

Additionally, we solved each of the test cases including the stochastic constraints. We conclude that using the stochastic constraints we are able to regulate the probability of lost demand. However, in terms of profit the stochastic constraints were not effective. We conclude that introducing buffer capacity, a common used inventory management technique, does reduce lost demand. We can obtain and maintain a desired service level by using a capacity buffer. However, these techniques only twist the prices to obtain a certain service level. In terms of profits, they are not beneficial. We saw those results for stochastic constraints on both the region level and the resource level.

Finally, we considered the computation times for each of the solved test cases. All of the test cases, for both the first and second step, were solved within 0.01 seconds. In our analytical comparison we concluded that the first step of the DRDP model is not expected to grow very large. Therefore, we do not expect that this step will lead to problems in terms of computation times. The second step of the DRDP model is very much comparable to the IRDP model. We expect that larger problem instances can be solved within reasonable time. Moreover, the IRDP model currently solves the actual problem instances (20+ million products) within several minutes.

8.1.1 Managerial Implications

In this report, we studied the competition between travel products to Greece, Spain, Turkey and Portugal. We found that the market shares of the popular regions (Greece and Spain) are harder to influence than the market shares of the less popular regions (Turkey and Portugal). We also found that the customers to Turkey and Portugal are willing to switch their destination easily if another region or the market is cheaper. Or alternatively, people are willing to go to Turkey and Portugal if the relative prices are favorable. Customers to Greece are only willing to switch to cheaper alternatives in Spain and the market to Greece. The customers to Spain are not willing to switch destinations, however when the market is cheaper they will switch to other trips to Spain in the market. These are insightful implications for managers, these findings can be used in order to steer marketing activities. Since customers to Turkey and Portugal are most price sensitive, we advise managers to invest in marketing activities for Turkey and Portugal.

We also concluded that the Dependent Region Dynamic Pricing (DRDP) model is expected to realize higher profits than the current IRDP model. We found that the DRDP on average led to 7.1% more profit than the IRDP model. We found that competition between regions and market information can add value to a tour operator's the pricing decisions. For managers, the question rises whether these improvements compensate the investments. In case of Sundio, the implementation of the model will require less effort. Since the DRDP model is an extension of the current IRDP model which is already implemented. We do not expect that the extended model will bring complications in terms of computation times. The data processing usually takes most of the time in the overnight process, the actual optimization is only a small step in terms of computation times.

We also advice to consider the fact that this model is dependent on the market information provided by GfK. Therefore, the reliability of GfK and the quality of the data must be studied in advance.

8.2 Further Research

In this section we will provide some ideas on directions of further research. First of all, in the DRDP model we fixed the number of market bookings over the optimization period. It might be interesting to replace these values by market demand estimates per time period. The market demand series can be used to determine the market demand estimates.

In this research we used the market bookings from GfK as an auxiliary data source. We showed that an auxiliary data source can add value to the pricing decisions. Since there are many more free data sources available it might be interesting for a tour operator to identify additional possible valuable information. Examples of this can be exchange rates, tax info, weather forecasts, popularity of destinations on social media, etc.

In the manegarial implications section (8.1.1), we concluded that the regression results provide insightfull results that can be used to steer marketing activities. In general, it can be Interesting to link marketing advertisement to operational information. This way, advertisements can be optimized on expected profit for the tour operator. Customers can be influenced to buy certain products that are beneficial for the tour operator. Recently, in cooperation with ORTEC, such algorithms are developed for Sundio.

In the airline and hotel industry, models are often extended to include the overbooking and cancellation of customers. The DRDP model can also be extended to include overbooking and cancellation of customers.

In the DRDP model, the demand is treated as given using point estimates. Because we treat the expected demand as given, the pricing decision from this model is only optimal if the realized demand is equal to the expected demand. We proposed a technique to control the lost demand probability in case of stochastic demand. This technique did not lead to higher profits. In order to obtain higher profits, it might be interesting to investigate methods from the hotel and airline industry in order to incorporate stochastic demand into the objective function.

Appendix I: Computational Results Regional Test Cases

	Base	IRDP	DRDP
Test Case 1: High, Equal, High			
Avg Profit (€)	373711	455969	546642
Std Profit (€)	54953	69104	83385
Profit % of IRDP	82.0 %	100.0 %	119.9 %
Profit % Highest	0.0 %	3.9 %	96.1 %
Profit % Lowest	73.4 %	25.8 %	0.8 %
Cap % Utilization	66.6 %	65.7 %	64.3 %
Avg Demand Lost	7.9	8.2	4.9
Test Case 2: High, Equal, Low			
Avg Profit (€)	373865	457846	547298
Std Profit (€)	25620	25871	35990
Profit % of IRDP	81.7 %	100.0 %	119.5 %
Profit % Highest	0.0 %	0.1 %	99.9 %
Profit % Lowest	84.1 %	15.9 %	0.0 %
Cap % Utilization	66.7 %	66.0 %	64.4 %
Avg Demand Lost	0.0	0.0	0.0
Test Case 3: High, Unequal, High			
Avg Profit (€)	380916	433875	557241
Std Profit (€)	87377	89212	115124
Profit % of IRDP	87.8 %	100.0 %	128.4 %
Profit % Highest	0.2 %	1.2 %	98.6 %
Profit % Lowest	73.9 %	25.3 %	0.8 %
Cap % Utilization	66.4 %	66.1 %	65.4 %
Avg Demand Lost	7.8	9.8	7.4

Table 8.1: Computational Results Test Case 1-3

	Base	**IRDP**	**DRDP**
Test Case 4: High, Unequal, Low			
Avg Profit (€)	382250	434072	557618
Std Profit (€)	35813	36663	46582
Profit % of IRDP	88.1 %	100.0 %	128.5 %
Profit % Highest	0.0 %	0.0 %	100.0 %
Profit % Lowest	84.0 %	16.0 %	0.0 %
Cap % Utilization	66.7 %	67.5 %	66.5 %
Avg Demand Lost	0.2	0.4	0.2
Test Case 5: Average, Equal, High			
Avg Profit (€)	363625	445149	534567
Std Profit (€)	45008	59201	71798
Profit % of IRDP	81.7 %	100.0 %	120.1 %
Profit % Highest	0.0 %	1.9 %	98.9 %
Profit % Lowest	86.3 %	13.1 %	0.6 %
Cap % Utilization	81.1 %	80.0 %	78.6 %
Avg Demand Lost	112.8	111.9	90.1
Test Case 6: Average, Equal, Low			
Avg Profit (€)	373617	457251	546971
Std Profit (€)	21538	27585	33174
Profit % of IRDP	81.7 %	100.0 %	119.6 %
Profit % Highest	0.0 %	0.0 %	100.0 %
Profit % Lowest	98.3 %	1.7 %	0.0 %
Cap % Utilization	83.3 %	82.3 %	80.4 %
Avg Demand Lost	3.2	4.7	1.4
Test Case 7: Average, Unequal, High			
Avg Profit (€)	370234	462223	542363
Std Profit (€)	52896	59085	75611
Profit % of IRDP	80.1 %	100.0 %	117.3 %
Profit % Highest	0.0 %	2.4 %	97.6 %
Profit % Lowest	87.7 %	11.4 %	0.9 %
Cap % Utilization	80.6 %	79.8 %	79.6 %
Avg Demand Lost	111.4	108.5	101.9
Test Case 8: Average, Unequal, Low			
Avg Profit (€)	381918	474489	556603
Std Profit (€)	34968	34810	46660
Profit % of IRDP	80.5 %	100.0 %	117.3 %
Profit % Highest	0.0 %	0.1 %	99.9 %
Profit % Lowest	99.3 %	0.7 %	0.0 %
Cap % Utilization	83.3 %	82.0 %	81.7 %
Avg Demand Lost	5.9	5.2	3.0

Table 8.2: Computational Results Test Case 4-8

	Base	IRDP	DRDP
Test Case 9: Low, Equal, High			
Avg Profit (€)	315405	395677	454093
Std Profit (€)	24274	32653	48797
Profit % of IRDP	79.7 %	100.0 %	114.8 %
Profit % Highest	0.1 %	5.9 %	94.0 %
Profit % Lowest	94.7 %	3.7 %	1.6 %
Cap % Utilization	93.7 %	92.8 %	90.1 %
Avg Demand Lost	592.7	558.3	377.2
Test Case 10: Low, Equal, Low			
Avg Profit (€)	333377	419677	488397
Std Profit (€)	5839	9915	11479
Profit % of IRDP	79.4 %	100.0 %	116.4 %
Profit % Highest	0.0 %	0.5 %	99.5 %
Profit % Lowest	100.0 %	0.0 %	0.0 %
Cap % Utilization	99.1 %	98.2 %	96.9 %
Avg Demand Lost	403.8	375.1	195.0
Test Case 11: Low, Unequal, High			
Avg Profit (€)	322212	402814	460416
Std Profit (€)	40034	42990	69039
Profit % of IRDP	80.0 %	100.0 %	114.3 %
Profit % Highest	0.0 %	3.4 %	96.6 %
Profit % Lowest	90.7 %	6.1 %	3.2 %
Cap % Utilization	93.7 %	92.7 %	89.0 %
Avg Demand Lost	589.6	562.2	376.9
Test Case 12: Low, Unequal, Low			
Avg Profit (€)	336083	421217	484715
Std Profit (€)	9661	10288	27659
Profit % of IRDP	79.8 %	100.0 %	115.1 %
Profit % Highest	0.0 %	0.4 %	99.6 %
Profit % Lowest	99.9 %	0.1 %	0.0 %
Cap % Utilization	97.7 %	96.9 %	93.6 %
Avg Demand Lost	448.7	417.2	214.2

Table 8.3: Computational Results Test Case 9-12

Appendix II: Computational Results Expanded Test Cases

	Base	**IRDP**	**DRDP**
Test Case 5.1: High, 3, 10			
Avg Profit (€)	186672	255414	272117
Std Profit (€)	7517	11306	11967
Profit % of Base	73.1 %	100.0 %	106.5 %
Profit % Highest	0.0 %	14.3 %	85.7 %
Profit % Lowest	70.4 %	21.5 %	8.1 %
Cap % Utilization	87.1 %	86.4 %	85.4 %
Avg Demand Lost	124.3	121.2	113.3
Test Case 5.2: High, 3, 100			
Avg Profit (€)	186238	256723	276121
Std Profit (€)	7488	11158	12015
Profit % of Base	72.5 %	100.0 %	107.6 %
Profit % Highest	0.0 %	14.7 %	85.3 %
Profit % Lowest	72.3 %	19.5 %	8.2 %
Cap % Utilization	85.8 %	86.3 %	85.6 %
Avg Demand Lost	128.1	122.3	111.1
Test Case 6.1: Low, 3, 10			
Avg Profit (€)	197374	267792	286773
Std Profit (€)	3442	5378	5925
Profit % of Base	73.7 %	100.0 %	107.1 %
Profit % Highest	0.0 %	9.0 %	91.0 %
Profit % Lowest	100.0 %	0.0 %	0.0 %
Cap % Utilization	93.7 %	92.4 %	92.2 %
Avg Demand Lost	33.2	29.9	24.3
Test Case 6.2: Low, 3, 100			
Avg Profit (€)	196111	267764	287244
Std Profit (€)	3460	5408	5870
Profit % of Base	73.2 %	100.0 %	107.3 %
Profit % Highest	0.0 %	10.4 %	89.6 %
Profit % Lowest	100.0 %	0.0 %	0.0 %
Cap % Utilization	93.6 %	92.6 %	92.4 %
Avg Demand Lost	35.2	31.1	27.3

Table 8.4: Computational Results Step1 and 2

Appendix III: Comparison Stochastic Demand Constraints

	DRDP $\theta = 0.3$	DRDP $\theta = 0.5$	DRDP $\theta = 0.7$
Test Case 1: High, Equal, High			
Avg Profit (€)	546078	546642	545624
Std Profit (€)	82224	83385	82225
Profit % Highest	33.2 %	33.5 %	33.3 %
Profit % Lowest	33.0 %	33.4 %	33.6 %
Cap % Utilization	64.2 %	64.3 %	64.2 %
Avg Demand Lost	5.3	4.9	5.6
Test Case 2: High, Equal, Low			
Avg Profit (€)	547456	547298	547324
Std Profit (€)	35786	35990	36305
Profit % Highest	33.4 %	33.3 %	33.2 %
Profit % Lowest	33.33 %	33.3 %	33.3 %
Cap % Utilization	66.5 %	64.4 %	66.4 %
Avg Demand Lost	0.0	0.0	0.0
Test Case 3: High, Unequal, High			
Avg Profit (€)	557651	557241	557511
Std Profit (€)	113573	115124	114216
Profit % Highest	33.9 %	32.7 %	33.4 %
Profit % Lowest	33.4 %	33.2 %	33.4 %
Cap % Utilization	66.5 %	65.4 %	66.4 %
Avg Demand Lost	6.9	7.4	8.4
Test Case 4: High, Unequal, Low			
Avg Profit (€)	557852	557618	557571
Std Profit (€)	45892	46582	46289
Profit % Highest	33.4 %	33.2 %	33.3 %
Profit % Lowest	33.5 %	33.4 %	33.3 %
Cap % Utilization	66.5 %	66.5 %	66.6 %
Avg Demand Lost	0.2	0.2	0.3
Test Case 5: Average, Equal, High			
Avg Profit (€)	444279	534567	535021
Std Profit (€)	71222	71798	71249
Profit % Highest	20.1 %	39.7 %	40.2 %
Profit % Lowest	79.9 %	9.9 %	10.2 %
Cap % Utilization	57.6 %	78.6 %	78.6 %
Avg Demand Lost	40.7	90.1	86.9
Test Case 6: Average, Equal, Low			
Avg Profit (€)	426774	546971	547623
Std Profit (€)	33199	33174	32908
Profit % Highest	12.4 %	43.4 %	44.2 %
Profit % Lowest	87.6 %	6.3 %	6.2 %
Cap % Utilization	61.4 %	80.4 %	79.7 %
Avg Demand Lost	0.5	1.4	1.3

Table 8.5: Stochastic Results Test Case 1-6

	DRDP $\theta = 0.3$	DRDP $\theta = 0.5$	DRDP $\theta = 0.7$
Test Case 7: Average, Unequal, High			
Avg Profit (€)	441239	542363	534938
Std Profit (€)	76189	75611	77261
Profit % Highest	19.2 %	40.7 %	40.1 %
Profit % Lowest	80.4 %	9.9 %	9.7 %
Cap % Utilization	60.1 %	79.6 %	78.9 %
Avg Demand Lost	56.2	101.9	107.5
Test Case 8: Average, Unequal, Low			
Avg Profit (€)	427213	556603	557665
Std Profit (€)	46215	46660	46187
Profit % Highest	13.5 %	43.3 %	43.2 %
Profit % Lowest	84.9 %	7.7 %	7.4 %
Cap % Utilization	59.2 %	81.7 %	81.3 %
Avg Demand Lost	0.8	3.0	3.2
Test Case 9: Low, Equal, High			
Avg Profit (€)	-	454093	453618
Std Profit (€)	-	48797	49029
Profit % Highest	-	52.2 %	47.8 %
Profit % Lowest	-	47.8 %	52.2 %
Cap % Utilization	-	90.1 %	89.9 %
Avg Demand Lost	-	356.6	370.0
Test Case 10: Low, Equal, Low			
Avg Profit (€)	-	488397	488544
Std Profit (€)	-	11479	11159
Profit % Highest	-	50.2 %	49.8 %
Profit % Lowest	-	49.8 %	50.2 %
Cap % Utilization	-	96.9 %	96.3 %
Avg Demand Lost	-	195.0	197.2
Test Case 11: Low, Unequal, High			
Avg Profit (€)	-	460416	460025
Std Profit (€)	-	69039	69203
Profit % Highest	-	51.1 %	48.9 %
Profit % Lowest	-	48.9 %	51.1 %
Cap % Utilization	-	89.0 %	90.2 %
Avg Demand Lost	-	376.9	368.8
Test Case 12: Low, Unequal, Low			
Avg Profit (€)	-	484715	484198
Std Profit (€)	-	27659	27588
Profit % Highest	-	50.4 %	49.6 %
Profit % Lowest	-	49.6 %	50.4 %
Cap % Utilization	-	93.6 %	93.3 %
Avg Demand Lost	-	214.2	212.8

Table 8.6: Stochastic Results Test Case 7-12

Appendix IV: Comparison Stochastic Demand Constraints Expanded Test Cases

	DRDP	DRDP	DRDP
	$\zeta = 0.3$	$\zeta = 0.5$	$\zeta = 0.7$
Test Case 5.1: High, 3, 10			
Avg Profit (€)	182073	272117	273521
Std Profit (€)	10726	11967	12142
Profit % Highest	3.9 %	48.0 %	48.1 %
Profit % Lowest	67.6 %	17.3 %	15.1 %
Cap % Utilization	82.1 %	85.4 %	85.8 %
Avg Demand Lost	70.3	113.3	112.0
Test Case 5.2: High, 3, 100			
Avg Profit (€)	180123	276121	276658
Std Profit (€)	10801	12015	12075
Profit % Highest	5.5 %	45.6 %	48.9 %
Profit % Lowest	65.1 %	17.7 %	17.2 %
Cap % Utilization	81.1 %	85.6 %	84.7 %
Avg Demand Lost	67.8	111.1	111.6
Test Case 6.1: Low, 3, 10			
Avg Profit (€)	184448	286773	284612
Std Profit (€)	4995	5925	5871
Profit % Highest	0.7 %	49.5 %	49.8 %
Profit % Lowest	96.7 %	1.9 %	1.4 %
Cap % Utilization	84.4 %	92.2 %	93.7 %
Avg Demand Lost	4.1	24.3	29.7
Test Case 6.2: Low, 3, 100			
Avg Profit (€)	186646	287244	287872
Std Profit (€)	5016	5870	5894
Profit % Highest	1.2 %	48.8 %	50.0 %
Profit % Lowest	96.4 %	2.0 %	1.6 %
Cap % Utilization	80.7 %	92.4 %	93.6 %
Avg Demand Lost	2.2	27.3	32.4

Table 8.7: Stochastic Results Test Case 5.1-6.2

Bibliography

[1] Allemand K., Fukuda K., Liebling T.M., Steiner E. "A polynomial case of unconstrained zero-one quadratic optimization", *Mathematical Programming Vol. 91(1)*, pp. 49-52, 2001.

[2] Axsäter S. "Inventory Control", *International Series in Operations Research & Management Science Vol. 90*, 2006.

[3] Aydin G., Ryan J.K. "Product line selection and pricing under the multinomial logit choice model", *Working Paper, Purdue University*, 2000.

[4] Bitran G., Monschein S. "Periodic Pricing of Seasonal Product in Retailing", *Management Science Vol. 43*, pp. 427-443, 1997.

[5] Bitran G., Caldentey R. "An Overview of Pricing Models for Revenue Management", *Manufacturing & Service Operations Management Vol. 5(3).*, pp.203-229, 2002.

[6] Bitran G.R., Caldentey R.A., Vial R. "Pricing policies for perishable products with demand substitution", *Working Paper, Stern School of Business, New York University*, 2007.

[7] de Boer S.V., Freling R., Piersma N., "Stochastic Programming for Multiple Leg Network Revenue Management", *European Journal of Operational Research Vol. 137*, pp. 72-92, 2002.

[8] Bollapragada S., Cheng H., Phillips M., Garbiras M. "NBC's Optimization Systems Increase Revenues and Productivity", *Interfaces Vol. 32(1)*, pp. 47-60, 2002.

[9] Boyd E.A., and Bilegan I.C., "Revenue Management and E-Commerce", *Management Science Vol. 49(10)*, pp. 1363-1386, 2003.

[10] Chintagunta P., Dube J.P., Singh V. "Balancing Profitability and Customer Welfare in a Supermarket Chain", *Quantitative Marketing and Economics Vol. 1(1)*, pp. 111, 2003.

[11] Deaton A., Muellbauer J. “Economics and Consumer behavior”, *Cambridge University Press*, 1980.

[12] De Mello M., Pack A., Sinclair M.T. “A system of equations model of UK tourism demand in neighbouring countries”, *Applied Economics Vol 34(4)*, pp.509-521 2002.

[13] Dong L., Kouvelis P., Tian Z. “Dynamic Pricing and Inventory Control of Substitute Products”, *Journal of Manufacturing Service Operations Management Vol 11(2)*, 2009.

[14] Elmaghraby W., Keskinocak P. “Dynamic Pricing in the Presence of Inventory Considerations: Research Overview, Current Practices and Future Directions”, *Management Science Vol. 49(10).*, pp.1287-1309, 2003.

[15] Fong N.M., Simester D.I. and Anderson E.T. “Private Label vs. National Brand Price Sensitivity: Evaluating Non-experimental Identification strategies”, *Working Paper MIT*, 2010.

[16] Geraghty M., Johnson E. “Revenue Management Saves National Car Rental”, *Interfaces Vol. 27(1)*, pp. 107-127, 1997.

[17] Granados N.F., Gupta A., Kauffman R.J., “IT-enabled transparent electronic markets: the case of the air travel industry”, *Information Systems and E-Business Management Vol.5(1)*, pp. 65-91 2007.

[18] Granados N.F., Gupta A., Kauffman R.J., “Information transparency in business-to-consumer markets: Concepts, framework and research agenda”, *Information Systems Research Vol.21(2)*, pp. 207-226 2010.

[19] Granados N.F., Gupta A., Kauffman R.J., “Online and offline demand and price elasticities: Evidence from the air travel industry”, *Information Systems Research*, forthcoming.

[20] Hanks R., Cross R., Noland R. “Discounting in the Hotel Industry: A new approach”, *Cornell Hotel and Restaurant Administration Quarterly Vol. 33(1)*, pp. 15-23, 1992.

[21] Hansen P., Meyer C. “Improved Compact Linearizations for the Unconstrained Quadratic 0-1 Minimization Problem”, *Discrete Applied Mathematics Vol. 157(6)*, pp. 1267-1290, 2009.

[22] Hawkins D. “Identification of Outliers”, *Chapman and Hall*, 1980.

[23] Heij C., de Boer P., Franses P.H., Kloek T., van Dijk H.K. “Econometric Methods with Applications in Business and Economics”, *Oxford University Press*, 2004.

[24] Hoch S.J., Kim B.D., Montgomery A.L., Rossi P.E. "Determinants of Store-Level Price Elasticity", *Journal of Marketing Research Vol. 32(1)*, pp. 17-29, 1995.

[25] Hormby S., Morisson J., Prashant D., Meyers M., Tensa T. "Marriot International Increases Revenue by Implementing a Group Pricing Optimizer", *Interfaces Vol. 40(1)*, pp. 47-57, 2010.

[26] Jenkins D. "Handbook of Airline Economics", *New York: McGraw-Hill*, 1995.

[27] Kimes S. "Revenue Management: A retrospective", *Cornell Hotel and Restaurant Administration Quarterly Vol. 44(4)*, pp. 131-138, 2003.

[28] Ladany S., Arbel A. "Optimal Cruise-Liner Passenger Cabin Pricing Policy", *European Journal of Operations Research Vol. 55*, pp. 136-147, 1991.

[29] Leeflang P.S.H., Parreño-Selva J. "cross category demand effects of price promotions", *Journal of The Academy of Marketing Research Vol. 2.* , pp.1-15, 2010.

[30] Liberti L. "Compact Linearization of Binary Quadratic Problems", *Published online in 4OR*, DOI:10.1007/s10288-006-0015-3, 2006.

[31] Loeve M. "Probability Theory", *Springer Verlag: Graduate Texts in Mathematics, Vol. 45(4)*, pp. 12, 1977.

[32] Malevergne Y., Pisarenko V., Sornette D. "On the power of generalized extreme value (GEV) and generalized Pareto distribution (GPD) estimators for empirical distributions of stock returns", *Applied Financial Economics, Vol. 16(3)*, pp. 271-289, 2006.

[33] Mandese J. "Taking Hint From the Travel World, ABC goes up against ritual: Aeronomics system expected to help troubled net set prices to fill seats.", *Advertising Age 11 May*, 1998.

[34] Mangion M., Durbarry R., Sinclair M.T., "Tourism competitiveness: Price and Quality", *Tourism Economics Vol. 11(1)*, pp. 45-68, 2005.

[35] Marriot J., Cross R. "Room at the Revenue inn", *Book of Management Wisdom: Classic Writings by Legendary Managers, New York: Wiley*, pp. 199-208, 2000.

[36] Massey F.J., "The Kolmogorov-Smirnov test for goodness of fit", *J. Am. Stat. Assoc. Vol. 46(253)*, pp. 68-78, 1951.

[37] Metters R. et al “The ’Killer Application’ of Revenue Management: Harrah’s Cherokee Casino & Hotel”, *Interfaces Vol. 38(3)*, pp. 161-175, 2008.

[38] Netessine S., Savin S., and Xiao W., “Revenue Management trough Dynamic Cross-Selling in E-Commerce Retailing”, *Operations Research Vol. 54(5)*, pp. 893, 2006.

[39] Pachon J., Iakovou E., Ip C. “Vehicle Fleet Planning in the Car Rental Industry”, *The Journal of Revenue & Pricing Management Vol. 5(3)*, pp. 221-236, 2006.

[40] Pak K., “Revenue Management: New Features and Models”, *ERIM Ph.D. Series Research in Management Vol. 61*, 2005.

[41] Patsouratis V., Frangouli Z., Anastasopoulos G., “Competition in tourism among the Mediterranean countries”, *Applied Economics Vol. 37(16)*, pp. 1865-1870, 2005.

[42] Perdue R.R., “Perishability,Yield Management, and Cross-Product Elasticity: A Case Study of Deep Discount Season Passes in the Colorado Ski Industry”, *Journal of Travel Research Vol. 41(15)*, 2002.

[43] Rao B.V., and Smith B.C., “Decision Support in Online Travel Retailing”, *Journal of Pricing & Revenue Management Vol. 5(1)*, pp. 72, 2006.

[44] Smith B., Leimkuhler J., Darrow R. “Yield Management at American Airlines”, *Interfaces Vol. 22(1)*, pp. 8-31, 1992.

[45] Song I., Chintagunta P.K. “Measuring Cross-Category Price Effects with Aggregate Store Data”, *Management Science Vol. 52(10)*, pp. 1594-1609, 2006.

[46] Talluri K.T., and van Ryzin G.J., “A Randomized Linear Programming Method for Computing Network Bid Prices”, *Transportation Science Vol. 33(2)*, pp. 207-216 1999.

[47] Talluri K.T., and van Ryzin G.J., “The Theory and Practice of Revenue Management”, *Kluwer Academic Publishers*, 2004.

[48] Tremblay P. “Pooling International Tourism in Western Europe”, *The Annals of Tourism Research Vol. 16(4).*, pp.477-491, 1989.

[49] Van-Kemp S. “Timeshare firm gets feet wet with the net”, *Adweek Magazines’ Technology Marketing Vol. 7*, 2004.

[50] de Vries G., Pak K. “Conquering the One-Size-Fits-All Disease: How to Put Revenue Management into Action”, *The Journal of Revenue & Pricing Management Vol 10*, pp.484-491, 2011.

[51] Williamson E.L., “Airline Network Seat Inventory Control: Methodologies and Revenue Impacts”, *Ph.D. Thesis, Flight Transportation Laboratory, Massachusetts Institute of Technology, Cambridge*, 1992.

[52] Zhang D., Cooper W.L. “Pricing Substitutable Flights in Airline Revenue Management”, *European Journal of Operational Research Vol. 197.*, pp.848-861, 2009.

[53] http://techcrunch.com/2011/01/03/j-p-morgan-global-e-commerce-revenue-to-grow-by-19-percent-in-2011-to-680b/