Integrated Pricing and Advertising at Sundio Group

Wim Jansen

UNIVERSITY OF TWENTE.

ORTEC
OPTIMIZE YOUR WORLD

Master Thesis Industrial Engineering and Management

Integrated Pricing and Advertising at Sundio Group

W. Jansen
2014, September

Public version

Educational institution
University of Twente

Faculty
Faculty of Behavioural, Management and Social sciences

Educational program
Master Industrial Engineering and Management
Track: Production and Logistics Management

Supervisory committee
Internal supervisors
Dr. M.C. van der Heijden
Dr. P.C. Schuur

External supervisor
Dr. K. Pak

UNIVERSITY OF TWENTE.

ORTEC
OPTIMIZE YOUR WORLD

Management Summary

At the moment Sundio uses two independent models for optimizing the prices of the holidays and the cost per clicks (CPCs) set for Google AdWords. Since the CPCs set in Google AdWords influence the demand, and for the revenue management model used for optimizing the prices a forecast has to be made for the demand, the effects of using the outcomes of both models interact and influence the profit made by Sundio. Therefore the next step is to set up a decision support model that integrates the pricing and marketing decisions in order to maximize profit.

The aim of this research is to come up with a possibility for modelling the integrated decision making for pricing and advertising at Sundio. Integrating both the results of changing prices and advertisement expenditures can result in finding new opportunities for increasing profit. The main research question is:

"What is an appropriate model for Sundio to support the decision making on both prices and advertisements in order to maximize profit?"

In this research we propose a mathematical model which optimizes the pricing and advertising decisions of Sundio. The main benefit of this model compared to the current model used is that it combines the effect of changing the price and changing the CPC in one model. Integrating all the decisions regarding Google AdWords results in improved benefits from Google AdWords since the CPCs are not optimized individually; the model looks for the best combination of CPCs and prices for all AdWords and product groups. Besides taking into account the capacity restrictions when optimizing the CPCs used in Google AdWords results in not spending money on attracting customers for packages which have (almost) no capacity left.

The proposed Integrated Pricing and Advertising model uses different parameters which have to be forecasted. We analyse different forecasting methods to investigate how these parameters should be forecasted. An interesting finding regarding forecasting the effects of the CPC set in AdWords is that there is a significant relation between the CPC set and the percentage that the advertisement is shown when someone searches on the related keywords. The current AdWords model does not take this relation into account.

For forecasting the different parameters we compared different forecasting methods. In all cases naïve forecasting methods result in a higher forecasting accuracy than the additive and multiplicative Holt Winter's method. We also investigated that by grouping the bookings based on the travel length of the holiday an accurate forecast can be made for the average number of PAX per booking. With the results of forecasting analyses we know how to forecast the different parameters used in the Integrated Pricing and Advertising model.

To test how the proposed Integrated Pricing and Advertising model performs compared to the models currently used we set up a simulated case study based on real life data. The main conclusion based on the simulated case study is that the Integrated Pricing and Advertising model performs better than the

current model used in almost every case; only for the less realistic scenario with unlimited capacity the proposed model performed a little worse than the current model used. Not taking into account the case with unlimited capacity, the improvements in the profit on the holiday packages (revenue – variable costs of packages) varied between 13% and 38%. Besides we concluded that using one small and one larges time period instead of only one time period also results in an average performance increase of 1 till 3%.

Based on the results of this study we come up with the following recommendations:

- Start with changing the current model used into a model using one small time period for the close future and one large time period for the remaining time. This change does not need a large investment but in the simulation study performed this change resulted in an increase in profit of 1 till 3% depending on the capacity available.
- Investigate how the effects of the CPC on the impression share can be taken into account in the current model used for Google AdWords. Not taking the effects into account could result in choosing a low CPC because the number of impressions is forecasted to high.
- Further analyse the possible effects of using the Integrated Pricing and Advertising model. The results of the simulated case study performed are interesting, however this is only based on one case study and therefore further research is needed. In the case study performed improvements of almost 30% in profit were noticed for a case with low capacity. For the case with high capacity an improvement in profit of almost 15% was achieved.

A possibility for further research in continuation on this research could be related to analyse how the proposed model performs for other cases. Besides it is interesting to investigate the possibilities for integrating other advertisement channels besides Google AdWords in the Integrated Pricing and Advertising model. Other possibilities for further research are to investigate how the model can be used in other industries like for example the car rental industry.

Preface

After studying for quite some years as a student at the University of Twente, in which I went on a Study trip to Peru and Chile, have been a full time board member Study Association Stress, played many games of (indoor) soccer, made a lot of friends and enjoyed a lot, it was time to finish my academic career. However, one more step needed to be taken: writing a master thesis report. By the time that you read this report, this final task has also been completed.

One year ago I started as an intern at ORTEC to perform a research about integrating the pricing and advertising models used by the Sundio Group. In the beginning there were big plans as for example taking into account both all the off line media and online media channels used by Sundio; during the project we better defined the scope of the project, but it still took a lot of time to perform this research. However, in the end I'm proud of the end result.

Without the support of other people I would not have been able to perform this research. In the first place I want to thank ORTEC, and especially Kevin Pak for giving me the possibility to perform this research and for all the support and feedback I got. The fact that when the end of my research was in sight I already started as a full time employee at ORTEC, confirms that I really enjoyed the time spent at ORTEC during my research. Also thanks to my colleagues at ORTEC for enabling me to perform my research within a very pleasant working environment and for answering a lot of questions about the complex system used at Sundio.

I also want to thank Matthieu van der Heijden and Peter Schuur for guiding me through the process of this master thesis. Your advice and comments have been most welcome. I appreciate the time and effort you put in supervising me during this research. Thanks also to Jasper Janssen for the feedback he gave me on this report.

Also thanks to my family for their support during this project, but also during my whole study career. And last but not least, thanks to my girlfriend Sarah Hollan for her support and patience in times that I needed it the most.

Wim Jansen

Contents

List of Tables

List of Figures

1 Introduction

This thesis is a report about the research performed at ORTEC. The research performed is the final assignment for finishing the Master study Industrial Engineering and Management at the University of Twente. The research is about how to integrate pricing and advertisement optimization at Sundio Group, a client of ORTEC.

In this chapter we give a short introduction of the research we performed. In Section 1.1 we give a general description about ORTEC. In Section 1.2 we introduce the online tour operator Sundio Group, the company where the research is about. We give a description of the problem faced by ORTEC and Sundio in Section 1.3.

1.1 Company description ORTEC

ORTEC is originally a Dutch company providing “advanced planning and optimization solutions and services” (ORTEC, 2013). Established in 1981 in Rotterdam, nowadays ORTEC has offices in Europe, North America, South America and the Pacific Region. Nowadays ORTEC has a broader view and focuses also on making informed, fact-based decisions based on thorough analysis by combining Operations Research, Business Knowledge and IT services. At ORTEC, almost 900 employees are working for around 1750 customers worldwide. These customers are operating many different industries; examples of well-known customers are Coca Cola, Shell, KLM, TNT, and Albert Heijn.

The research takes place at the department Living Data. This department focuses on value creation through innovation and seizing marketing/ecommerce opportunities by using statistical analyses of data. A large customer cooperating with this department is Sundio Group. In the next section we give a description of this company since the project is related to this customer.

1.2 Company description Sundio Group

Sundio Group (further: Sundio) is the largest online tour operator in the Netherlands and arranges holiday packages for over 800,000 people yearly (Sundio Group, 2013b). Currently, Sundio has over 170 employees at the headquarters in Rotterdam. Besides they have offices in Belgium, Germany, France, Denmark, the United Kingdom, Poland and Spain. At the different travel locations they have around 350 travel guides and hostesses.

Sundio operates under different brand names; those brand names are Sunweb, Sudtours, Jiba, GOGO, Husk Studentenreizen, X-travel, Eliza was here, Skikot, SkiHorizon and TravelHorizon (Sundio Group, 2013a). By taking over small tour operators, Sundio extends the number of holiday packages they can offer to customers.

As Sundio is operating in a transparent online market it is important that when a customer is looking for a holiday offer, he is offered the right price. Considering that the products of Sundio have a perishable and fixed capacity, it is important for Sundio to match demand with supply. When demand is low compared to capacity, Sundio can choose to invest more in advertisement or to lower the prices offered to customers to attract more customers. When demand is high compared to the capacity it is the other way around.

Considering that Sundio is a commercial company, they want to have their pricing and advertisement expenses set in such a way that profit is maximized. Therefore Sundio cooperates with ORTEC in order to optimize the pricing and advertisement decisions made. In Subsection 1.2.1 and Subsection 1.2.2 we give a short explanation of how Sundio makes pricing and advertisement decisions.

1.2.1 Pricing at Sundio

In cooperation with Sundio, ORTEC created a tailored model to combine classic pricing techniques with Revenue Management (RM). An important difference compared to traditional airline RM models is that Sundio uses one price for a product at a time, instead of multiple prices being available for one product at the same time (De Vries & Pak, 2011). This means that the model focuses on setting the single price used for a product at different time periods, instead of assigning a number of PAX to different classes.

Figure 1 illustrates the price development when using capacity control (traditional airline RM models) or dynamic pricing (as used at Sundio). When using capacity control multiple prices for different classes are available at the same time. At certain moments in time, when the booking limit for a class is reached, a class is not available anymore. In the case of dynamic pricing, as in the model of ORTEC, one price is available at a certain time interval. The prices used and the intervals for these prices influence the overall demand. In Figure 1 we see that for the dynamic pricing at Sundio, the price is decreasing over time. Sundio has chosen to use the highest price first based on the assumption that people booking their holiday a long time in advance are looking for certainty and are willing to pay for that. This price development is in line with the prevailing market, prices of competitors also seem to decrease over time. This is the opposite of what happens in the airline industry where mainly business travellers have to pay a premium for their late bookings.

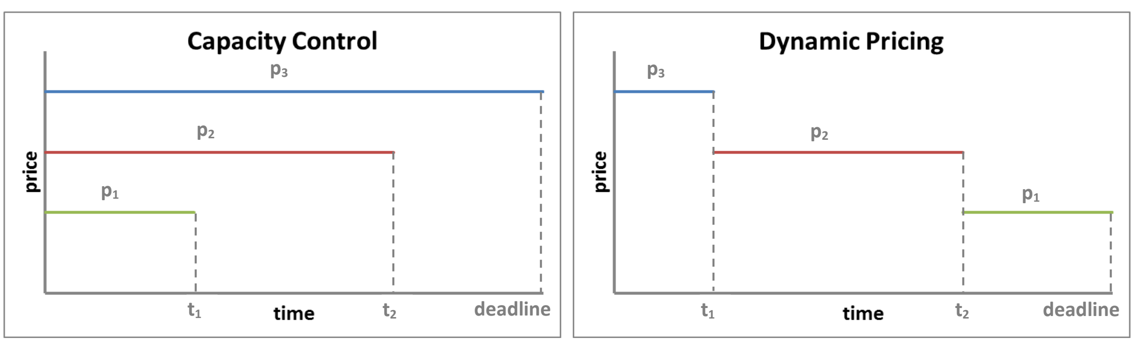

Figure 1: Availability of different prices when using capacity control or dynamic pricing

Based on the pricing model, an RM system is developed to support the Yield department at Sundio. This system, called GoGoYield (GGY), estimates arrival rates of customers by combining the price elasticity of every product with an estimated arrival rate at a reference price. Based on these estimated arrival rates and the time left before departure, the system advises which prices to use to maximize the expected profit. In this way, over 75 million prices are optimized on a daily basis. In Section 3.2 we go into further depth about this pricing system.

1.2.2 Advertisement at Sundio

On a yearly base Sundio spends around € 20 million on advertisements. The marketing activities at Sundio are divided into advertisements with a focus on direct sales and advertisements for branding. Examples of advertisements with a focus on direct sales are newsletters and Google AdWords[1] advertisements; these are directly related to attracting more people to the websites. Branding activities are for instance TV/radio commercials, and advertisements in magazines; these activities focus on long term results.

For that part of the marketing with a focus on direct sales, ORTEC created a system called GoGoMarketing (GGM). This system offers an algorithm for the selection of packages to show as a daily offer, tip, or high light at the website or in the newsletter. Before ORTEC implemented GGM at Sundio, the marketing department was mostly focused on holiday packages with the highest remaining capacity, while this does not guarantee that a higher profit is generated. For example promoting a package with a higher margin likely results in a higher extra profit compared to promoting a holiday package with a high capacity left, but a low margin. GGM is also used to determine optimal bid amounts for keyword advertising via Google AdWords[2].

For the branding activities a study performed by ORTEC is used by Sundio, but decisions made by the brand managers at Sundio are for a large part still based upon expert opinion. Decisions made by the brand manager are for example the size of the budget, how the budget is allocated over the different media available, and at what point in time it should be spent.

1.3 Problem description

At the moment the decisions made by the marketing department (partly based on GGM) and the prices set by GGY are independently optimized. Since they both interact and influence the profit made by Sundio, the next step is to set up a decision support model that integrates the price and marketing decisions in order to maximize profit.

By combining information about the effects of different methods of advertising with a forecast of how the conversion rate (number of bookings versus the number of website visitors) changes by changing the price, it should be possible to calculate the ideal advertising mix combined with the right price. For example when a product has a low demand compared to the capacity available, Sundio could lower the price or increase the expenditures on advertisement in order to increase demand. Besides it could also be profitable to increase expenditures on advertisement and at the same time increase the prices of different products.

---

[1] Google AdWords is an advertisement program which "enables any advertiser to purchase individualized and affordable keyword advertising that appears instantly on the google.com search results page." (Google, 2000) By using this program, an advertiser can show advertisements above and next to the search results.

[2] Google assigns advertisement positions based on the bids made for a key word; an optimal bid amount is generated based on the expected costs (clicks times bid amount) and the expected revenues (conversions times margin per sale) for different positions.

Currently, for the revenue management system, a forecast is made for the demand as a whole. We expect that by dividing the demand function into an attraction function and a conversion function, better forecasts of the demand can be made. Besides it is easier to integrate the results of changes in pricing and advertising decisions, since the pricing decisions influence the conversion function and advertising decisions influence the attraction function. Figure 2 gives a schematic overview of this concept. First potential customers should be attracted to the website by using different types of advertising. When a customer is at the website he should be converted to a sale. The conversion rate depends on things like price, availability, product mix and interests of the potential customers attracted to the website.

Figure 2: Schematic overview of generating profit by attracting customers to and convert customers at the website

The main problem investigated in this research is about how to model the interaction between the pricing and advertising decisions. We focus on the online advertisement via Google AdWords since there is a lot of data available related to this advertisement method. Therefore, compared to advertisements on TV or radio commercials, it is easier to relate the results of advertisements to the decisions made.

2 Research Design

In the previous chapter we discussed the problem faced and the different companies involved in this research. In this chapter we elaborate on the research design. In Section 2.1 we present the goal of the research and the related deliverables. We describe the research questions in Section 2.2. In Section 2.3 we discuss the scope of the research and in Section 2.4 how we collect the data. We give a definition of the most important terms used in this thesis in Section 2.5. In Section 2.6 we give an explanation on the outline of the thesis.

2.1 Research goal and deliverables

The goal of the research is to increase the profit of Sundio by supporting better decision making on both pricing and advertising. To fulfil this goal we have to come up with a possibility for mathematically modelling the integrated decision making for pricing and advertising. For this research advertisement decisions are restricted to the decisions regarding at what point in time, for how long and for which price.

The final deliverable of this research is a report describing how the pricing and advertising decisions can be modelled and optimized. Besides we investigate whether the new model provided outperforms the current model. The report contains an explanation of the data analysis performed, a clarification of the model, possibilities for future research and recommendations related to the research.

2.2 Research questions

To achieve the research goal we have to answer the following research question:

"What is an appropriate model for Sundio to support the decision making on both prices and advertisements in order to maximize profit?"

Furthermore we have formulated a series of stepwise sub-questions in order to explore the different aspects of the main research question.

1. *What models and systems are currently used regarding pricing and advertising at Sundio?*

After answering the first question we know which models and systems are currently used at Sundio to substantiate the pricing and advertising decisions. These models and systems can be used as a first input in constructing a new model taking into account the relation between marketing and pricing.

2. *What literature is available related to the decision making in advertising and revenue management?*

The answer of the second question is used to get more insight in the advertising and revenue management issues related to the research. In the literature we search for researches performed related to revenue management and decision making for online advertisement. Besides we discuss different forecasting methods which can be used when setting values for the parameters used in a new model.

3. *How can the effects of the decisions in pricing and advertising be combined in an integrated model?*

When answering the third sub-question we come up with a model which combines the effects of the pricing and marketing activities at Sundio. This model should be able to predict the effects of different marketing and pricing decisions.

4. *How can we use the data available regarding pricing and advertising decisions to forecast the effects of pricing and advertising on attraction and conversion?*

After answering sub-question 4 we have more insights regarding the forecasting of different parameters related to the model proposed when answering sub-question three; we can use this information in determining the optimal pricing and advertising strategy of Sundio.

5. *How does the model proposed perform compared to the current method used?*

After formulating the model we have to test the practical usage of the model. After answering this question we know if the final model can be used to optimise the decisions made regarding advertising and pricing. If so, we determine what the additional benefits are of using the model compared to the current model and whether the expected benefits are interesting for practical usage.

2.3 Research scope

In this section we give an outline of what the research covers. The research is focused on developing a combined pricing and advertising model for Sundio, however it could be possible that the result is also applicable to other online tour operators. To confine the amount of data used in the analysis, we focus on one brand of Sundio. We choose to only use the data related to Sunweb, since this is the brand with the highest impact on the results of Sundio.

A focus regarding the marketing activities in this research is related to pulse and step marketing actions. Hanssens, Parsons, and Schultz (2003) distinguish pulse and step actions in marketing as follows: a pulse action can be started and stopped at any moment (e.g., advertisements in a magazine), while step actions have a long-lasting nature (e.g., offering a new holiday package). In our research we focus on the pulse actions because the model should be able to support the decisions related to the pulse actions.

In this research we also chose to focus on quantity instead of quality regarding the advertisement used. We focus on the marketing decisions regarding the budgets and expenditures, not on for example the text used within. Since it is difficult to identify the results of off line advertisement we chose to focus on online advertisement. We concentrate on Google AdWords since this is the main online advertisement medium used by Sundio and there is a lot of data available related to the historical decisions and performance of Google AdWords.

2.4 Data collection

In this section we describe what data we use and how we are able to get this data. Since ORTEC and Sundio cooperate in generating a well working pricing and marketing system we are able to get a lot of data directly from Sundio. Sundio has an extensive database with information about the number of passengers, the revenue, costs, destination, booking date, and more for every booking. This information can be retrieved by executing queries in the MySQL database.

From the data warehouse we are also able to gather information related to the bookings of competitors of Sundio. Sundio retrieves this information from the market research institute GfK. The database contains information about for example the booking date, the departure date, and the revenue of bookings made at competitors of Sundio. By using this data we can compare the results of Sundio with the results of the market and competitors.

We also have access to the Google AdWords accounts used by Sundio. Via the web interface of Google AdWords we can gather information about for example the number of clicks on an advertisement, the position of an advertisement and the cost per click. Unfortunately Google does not keep track of the maximum bids set in the past. Besides we have access to the Google Analytics account used by Sundio. Google Analytics is a service of Google which keeps track of the visitors at a website. By using Google Analytics is it possible to see how many people visited the website of Sundio and which pages they visited.

In our research we have to combine the data from the different sources in order to set up a new model regarding the advertisement and pricing decisions made at Sundio. By using advanced statistical methods like regression analysis we search for correlations which we can use in our model.

2.5 Definition of terms

For a clear understanding of the terms used in this research we give a definition of jargon used in this report.

Accommodation group — An accommodation group is a group of accommodations with the same rating, room type and region. An accommodation group is used to aggregate packages in a product group. Packages with the same departure date, travel length and accommodation group are grouped for optimization.

Booking — A booking is a reservation made for a package. One booking can result in multiple PAX. A booking can be cancelled; we ignore this fact in our research since the cancellations are not part of the data used in this research.

Conversion — In the context of Sunweb we call a visitor of the website who makes a booking a conversion.

Conversion rate — A conversion rate is the number of conversions divided by the number of website visitors. For example when 1000 people visit the website of Sunweb in a specific week, and the number of bookings in that week is 3, the conversion rate is 0.3%.

Guaranteed capacity	Guaranteed capacity is capacity for which Sundio has to pay costs regardless of the capacity being used or not. The costs of guaranteed capacity can be seen as fixed costs.
Market	The market is defined as all tour operator activities in the Netherlands. For example when a tour operator is operating in different countries, we only take into account their activities in the Netherlands since these can influence the results of Sundio.
Non-guaranteed capacity	Non-guaranteed capacity is the opposite of guaranteed capacity. Sundio only has to pay the costs for non-guaranteed capacity when it is sold. Therefore the costs of non-guaranteed capacity can be handled as variable costs.
Package	A package is a combination of a transport to and from a location and an accommodation. A package contains the following properties: used brand (Sundio offers holidays via different brands), destination, accommodation, room type, transport type, start date, outbound flight, end date, and inbound flight. Multiple packages can have the same properties.
PAX	A PAX is a single passenger for which 1 unit of capacity is reserved.
Product	A product is a distinct package as it is offered for sale that embodies all attributes associated with the package. For example a product could be the trip offered by Sunweb from Amsterdam to Side (Turkey) on the 6th of October, a seven day stay in the Hotel Terrace Elite Resort and a return flight to Amsterdam. For this product multiple packages can be available, so this product can be booked by different customers.
Product group	A product group contains multiple products with shared characteristics. For example at destination country level (the different levels used at Sundio are described in Section 3.1) a product group could contain all products with destination country Turkey, while a product group at transport type level could contain all products to “Turkish Riviera (Turkey)” bookable via Sunweb and having transport type “flight”.

2.6 Outline thesis

This section contains a short outline of the different chapters in this thesis. In Chapter 1 we gave a short introduction to the problem. In this chapter we described the research design. In Chapter 3 we describe the current situation. The current situation describes the forecasting methods, pricing model and advertisement model used at Sundio. After describing the current situation we discuss some opportunities for improvement.

In Chapter 4 we give a short overview of the developments in Revenue Management based on a literature research performed. Besides we discuss several forecasting methods found in the literature.

In Chapter 5 we propose a new integrated model which optimizes the prices set for different product groups and the CPC to use in Google AdWords. We start with discussing a simple model which only takes into account one product group with no capacity restriction and one AdWords group. Within this chapter we expand the model to take into account multiple product groups with related capacity restrictions and multiple AdWords groups.

In Chapter 6 we evaluate the performance of the forecasting methods discussed in Chapter 4 in order to give an advice regarding which forecasting methods to use for forecasting the values of the different parameters used in the Integrated Pricing and Advertising model. In Chapter 7 we discuss the simulation study performed to evaluate the model proposed. In this chapter we compare the proposed model with the current model used based on results of our simulation study.

We give an overall conclusion of the research in Chapter 8. In this chapter we also discuss the limitations of our research and we mention some opportunities for further research.

3 Current Situation

In this chapter we answer the first sub-question of our research: "What is the current situation regarding pricing and advertising at Sundio?" In Subsection 1.2.1 and Subsection 1.2.2, we already gave a brief description of the current situation. In Section 3.1 we discuss the forecasting methods used and in Section 3.2 the pricing model used at Sundio. The information in these sections is based on (i) internal documentation of ORTEC, (ii) interviews with people working on the Sundio project at ORTEC, and (iii) De Vries and Pak (2011), which describes the dynamic pricing model used at Sundio. In Section 3.3 we describe the current practise in advertising. After explaining the current practices at Sundio; we discuss some opportunities for improvement in Section 3.4. We outline the main findings and summarize our answer on the first sub-question of our research in Section 3.5.

3.1 Forecasting at Sundio

Sundio uses a data warehouse in which they store amongst others data related to the bookings made. This data is important for the forecasting at Sundio since the historical data is used to make predictions about the price elasticity and the expected demand. In this section we explain how a forecast is made for the demand function. The demand function indicates how many PAX are expected in the remaining booking period; further in this section we go into more depth about the demand function. This demand function is used to generate input for the pricing model which we describe in Section 3.2. Figure 3 displays the different levels of product groups used at Sundio. A forecast is made for the demand function of a product group at level 6 for every departure day and travel length.

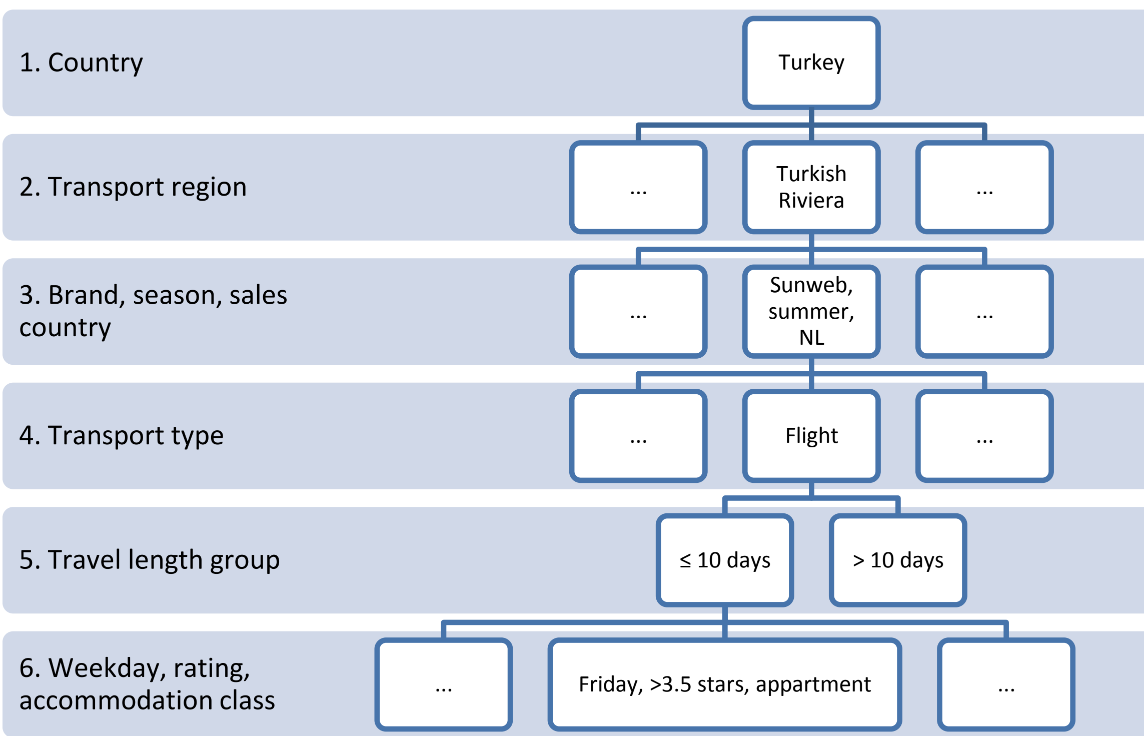

Figure 3: Product group tree

Figure 4 shows a summary of the different steps for forecasting the demand function at Sundio. We discuss the first five steps in more detail in Appendix A; the final step is explained in this section.

The first step in making a forecast at Sundio is to determine a booking pattern. This booking pattern is related to the average price development in the previous two years. Based on this booking pattern an initial forecast is made for the remaining number of PAX for a product group. This initial forecast is smoothed over different departure weeks. The next step in forecasting the number of PAX is applying a correction factor based on the load factor (PAX/capacity). Besides a price elasticity is calculated. The price elasticity and the expected number of PAX are combined to determine the demand function.

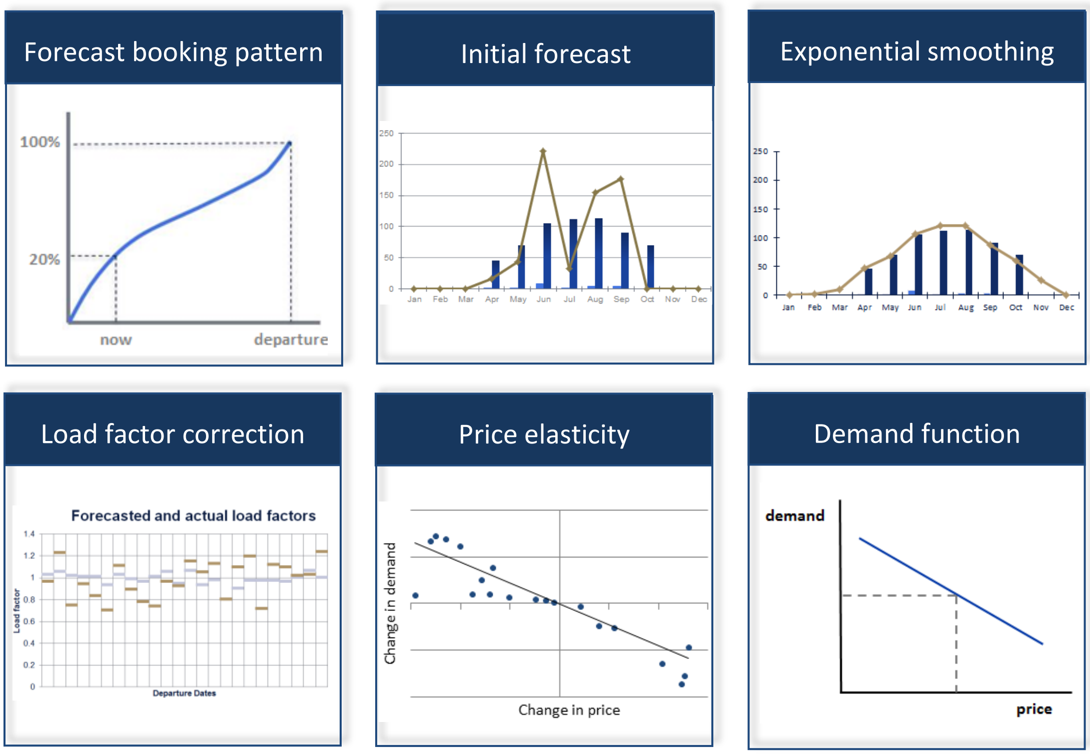

Figure 4: Simplification of the different steps in forecasting the demand function

Now that a forecast is made for the price elasticity and the expected remaining demand, a demand function can be determined. The demand function is based on the number of PAX expected in the remaining booking period, the reference price which based on the average historical price changes and the price elasticity. The model used at Sundio takes five possible prices for the next period into account related to a price change of 0%, 5% or 10%. These price changes can be multiplied with the price elasticity in order to forecast the change in remaining demand related to the price change. We illustrate these calculations in Example 3.1.

Example 3.1. *For our example product (see Appendix A) we know that the current price of our product group is € 732.75 and the average historical price change in the previous two year is -0.368%. Therefore the reference price used for calculating the demand function of our example product is € 732.75*(1-*

0.00368)=€ 730.07. This means that when we would use a price of € 730.07, we forecast 165.86 PAX to be booked in the remaining booking period. For this prediction is assumed that the price in the remaining period is always equal to the reference price calculated.

The price elasticity was calculated to be -2.00, so when the price is lowered with 5%, the expected remaining PAX would be 165.86-0.05-2.00*165.86=182.45. In the same way we can calculate the expected remaining PAX for other prices, the resulting demand function is displayed in Figure 5.*

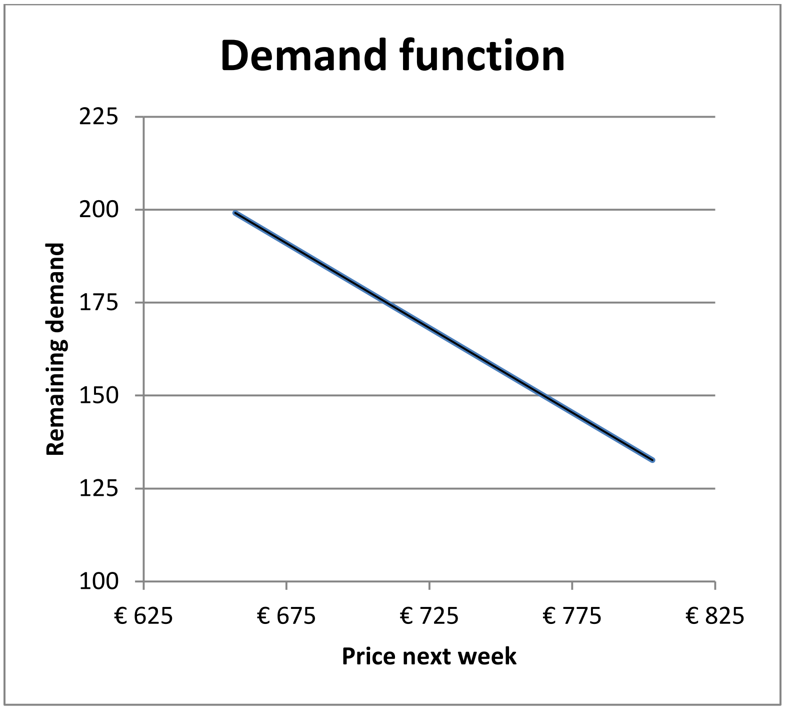

Figure 5: Demand curve example product

The price used in the demand function is not the price during the remaining booking period, it is the price used for the next week taking into account the average price developments in the previous two years for the weeks thereafter.

We assume that the cost price for our example product is € 450. Based on this cost price, the capacity of 220 and the demand curve we can calculate the revenue, costs and profit we expect when using different prices. These values are shown in Table 1. The first column shows the change in price relative to the price last week. The second column shows the related price. Column 3 shows the relative divergence compared to the reference price. This difference results in a relative difference in the expected demand as shown in column 4. The next column shows the expected remaining demand. Note that since currently there are already 18 PAX booked and the capacity is 220, the expected remaining PAX cannot be more than 220-18=202. The last three columns contain respectively the expected revenue, costs and profit over the remaining bookings.

Table 1: Forecast revenue, costs and profit for example product

Price change	New price	Difference reference price	Relative demand difference	Expected remaining PAX	Revenue	Variable costs	Profit
-10%	€ 659.48	-9.7%	33%	198.01	€ 130,583	€ 89,104	**€ 41,478**
-5%	€ 696.11	-4.7%	16%	181.32	€ 126,222	€ 81,596	**€ 44,626**
0%	€ 732.75	0.4%	-1%	164.64	€ 120,640	€ 74,088	**€ 46,552**
5%	*€ 769.39*	*5.4%*	*-18%*	*147.95*	*€ 113,834*	*€ 66,579*	***€ 47,255***
10%	€ 806.03	10.4%	-35%	131.27	€ 105,806	€ 59,071	**€ 46,735**

We have also displayed the expected revenue cost and profit of our example product when using different prices in Figure 6. We see that the highest expected profit can be achieved by increasing the price for the next week with 5% to a price of € 769.39. Note that in this case the network of capacity constraints (see Section 3.2) is not taken into account. For example it could be that there are many bookings for a package with the same destination but another accommodation rate resulting in a lower capacity for the departure flight. Besides we assumed that the complete capacity is non-guaranteed capacity. When the capacity would be completely guaranteed it could be better to use a lower price since Sundio has to pay for the seats even if they are not used.

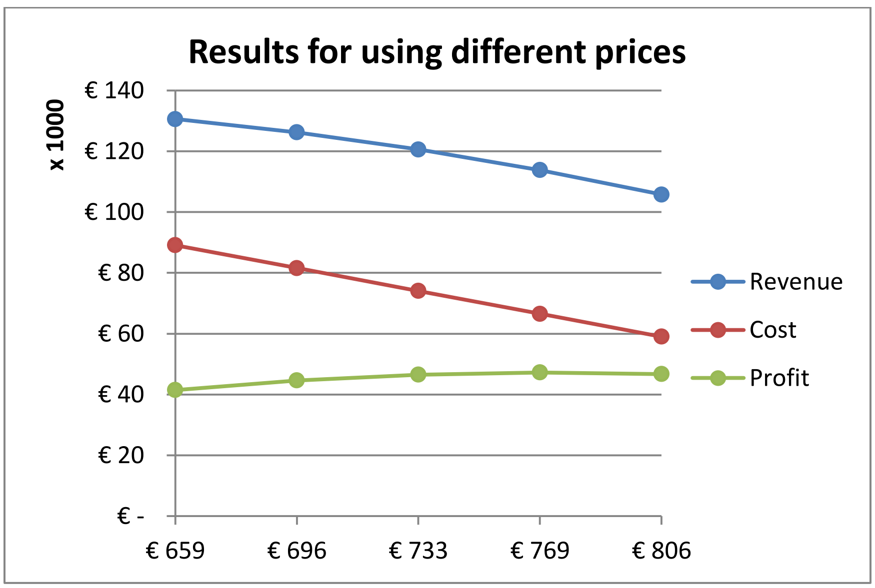

Figure 6: Expected revenue, variable cost and profit for the example product for different new prices

An advantage of the system used is that historical patterns are taken into account and it quickly responds to the current state of bookings; besides different forecasts can be made per product and departure date. By using a smoothing technique for the expected number of PAX booked, sensitivities in the start-up period and the individuality of the forecast are taken away.

We also noted some objections to the forecasting system used. The forecasting system is based on the prices of the bookings made while these depend on for example the availability of the apartments.

Another shortcoming in the forecasting method could be that price elasticities for products with a different departure date are equal over time and independent on the booking time.

3.2 Pricing model at Sundio

In Section 3.1 we explained how a forecast of the demand curve is made. Based on this demand curve, the cost price, and the capacity we can make a forecast of the profit for using different prices. However since multiple products can use the same resource, the calculation of the profit is more complex because capacity restrictions can overlap. To find an optimal solution taking into account this network of capacity restrictions an LP model is solved. In this section we describe the model used.

In order to reduce the number of restrictions and variables used, the model uses not individual products, but products groups. The product groups have the same characteristics as the product groups used in the forecasting stage, however product groups used in this model also differentiate based on the location, departure city, departure date and travel duration. By also differentiating on these characteristics products making use of different transport resources are not grouped together. The forecasted demand for the different product groups is divided over the product groups used for the optimization based on the number of PAX booked. We illustrate the difference between the product groups used for forecasting and the product groups used for optimization in Example 3.2.

Example 3.2. *In Example 3.1 we set up a demand function for our product group with destination region Turkish Riviera on a Friday in week 36 and a duration of smaller or equal than 10 days in a five star hotel booked via Sunweb. We assume that there is only one departure city available for this product group and that the different products in the product group have a length of 7, 8, 9 or 10 days.*

*We calculated that when we would not change the price for the next week, 165.64 PAX would be booked in the remaining booking period. These 165.64 PAX would then be divided over the different product groups used for the optimization. So if we assume that for the product groups with a duration of 7, 8, 9 and 10 days, the number of PAX booked at the 18th of February is respectively 3, 5, 7, 3 (18 in total). Then for the optimization process we expect the remaining demand for the product group with a duration of 8 days to be 165.64/18*5=46.*

For describing the mathematical pricing model used at Sundio we first introduce the sets, parameters and decision variables used.

Indices:

i	Index of product groups
j	Index of price classes
k	Index of resources

For every product group a set of prices is defined which contains the prices possible to assign to a product group. These prices are the current price used, and the prices 5% or 10% above and below the current price. The model also uses different resources which can be a transport or an (set of) accommodation(s) at a specific date.

Parameters:

p_{ij} Selling price of product group i in price class j

q_i Sum of variable flight and accommodation costs of product group i

r_{ij} Number of PAX expected in the remaining booking period for product group i when using price class j

c_k Units of capacity available for resource k

The selling prices and the cost prices are used to calculate the profit margin in the objective value. For the cost prices the variable costs are taken, therefore the system does not need to differentiate the products based on whether seats or accommodations are guaranteed or not (guaranteed seats/accommodation just have zero costs). The units of capacity available can be the number of seats available for a specific transport or the number of rooms at an accommodation.

Sets:

S_k Set of product groups using resource k

Since multiple products make use of the same transport or the same accommodation, sets are defined for all the products using a specific transport or accommodation. By using these sets the network effects of the different packages are taken into account. In Figure 7 we see an overview of the different network effects involved in the decision making at Sundio. Different products could share the same accommodation or make use of the same flight. Therefore for every flight and for every night at an accommodation a capacity constraint is used. The example product mentioned in the previous section could make use of nine resources: the flight to the Turkish Riviera at the 13th of August, the flight from the Turkish Riviera at the 20th of August, and seven nights at the accommodation where every separate night is defined as a specific resource.

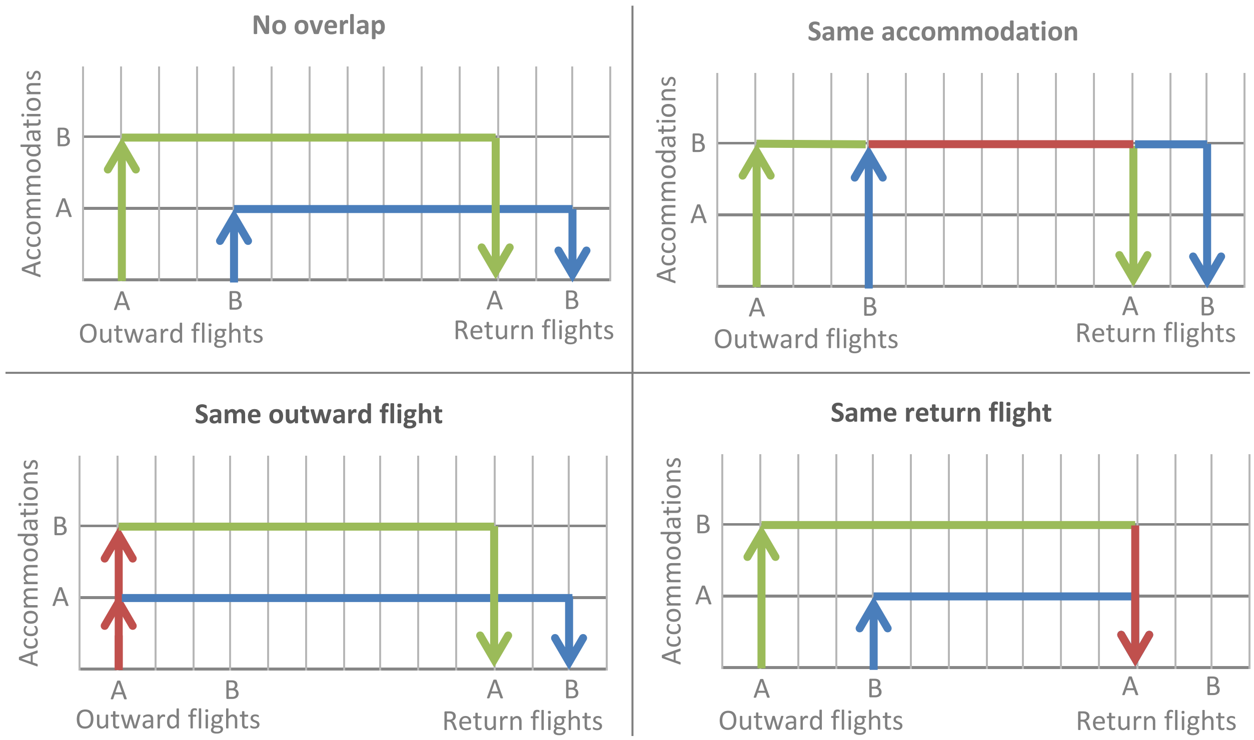

Figure 7: Illustration of the different network effects involved in the model of Sundio. A horizontal line represents a stay at an accommodation, a vertical arrow represent a transport.

Decision variables:

X_{ij} Fraction of the remaining booking period for which price class j is allocated to product i

The decision variable used indicates the fraction of time of the remaining booking period that a product should be offered for the price in price class j. For example when $X_{1,1} = .3$, $X_{1,2} = .7$, and $X_{1,3} = X_{1,4} = X_{1,5} = 0$ and the remaining booking period is 100 days, the products in product group 1 should be offered 30 days for price class 1 (for example increase 10%) and 70 days for $price\ class\ 2$ (for example increase 5%) .

Now that we have explained which indices, parameters, sets and decision variables are used, we can discuss the model used for the optimisation of the prices. The linear programming model used at Sundio is formulated as in (3.1)-(3.4).

Maximize:

$$Z = \sum_i \sum_j (p_{ij} - q_i) r_{ij} X_{ij} \tag{3.1}$$

Subject to:

$$\sum_{i \in S_k} \sum_j r_{ij} X_{ij} \leq c_k, \qquad \forall\, k \tag{3.2}$$

$$\sum_{j} X_{ij} \leq 1\,, \qquad \forall\, i \tag{3.3}$$

$$0 \leq X_{ij} \leq 1, \qquad \forall\, i,j \tag{3.4}$$

The objective function in (3.1) is used to maximize the total margin based on the expected arrival rates for different prices and the fractions that we use a specific price. For guaranteed capacity the costs are handled as fixed costs, for non-guaranteed capacity the costs are considered variable costs. When only a part of the remaining capacity of a product is guaranteed capacity, the total capacity is considered as guaranteed capacity. The margin per product is calculated by subtracting the variable costs from the price used at the website. Note that the objective value is not the total profit of Sundio since for some accommodations or transports a fixed capacity is bought; these costs are not taken into account in the objective function.

Constraint (3.2) is a capacity constraint; the total demand for a capacity cannot exceed the capacity available. The model determines the total demand for a capacity by taking the sum of the forecasted arrival rates multiplied with the fractions of time allocated. Constraint (3.3) and (3.4) are used to set restrictions to the sum of the fractions of a time period allocated to a product. Constraint (3.3) prevents that more than 100% of the time is assigned; constraint (3.4) prevents that negative fractions of time are assigned. Note that as long as the profit margin of a product is more than zero, the sum of the fractions of time assigned to that product is 1 due to maximization of the margins.

To bring the model into practice, Sundio made some changes in how they implemented the model in GGY. An example is that, because the model runs once every night, they only use the highest price class if X_{ij} multiplied with the remaining booking period is more than one day. For example when the remaining booking period is 30 days, and the highest price has a ratio of .01 (equal to one thirtieth of a day) and the second highest price a ratio of .2 (equal to 6 days), the price of the product for the next day is set at the second highest price.

The size of the model depends on multiple decision variables; the number of packages, capacities, and price classes per product all influence the total number of constraints. As mentioned before Sundio limits the number of possible prices by taking prices close to the previous price used. Besides similar products are grouped in a product group and optimized as one product, a result of this is that similar products have the same price change. Every night over 75 million prices are optimized by the system. These prices are uploaded to the websites of Sundio automatically.

3.3 Advertising at Sundio

In this section we describe the different types of advertisement used at Sundio. As explained in Subsection 1.2.2, Sundio makes a distinction between branding advertisement and direct sales advertisement. At Sundio, the brand manager decides in cooperation with the marketing manager what the marketing budget is. The ratio between branding and direct sales advertisement differs per brand; if a brand operates in a niche, more budget is spent on direct sales and less on creating brand awareness. In this research the scope is limited to the brand Sunweb.

Sunweb makes use of different possibilities for branding advertisement. The branding activities are important to let the direct sales advertisement be effective. The brand manager and the marketing manager decide which activities are used for a brand. The decisions are made on the expert opinions of both. The brand manager is responsible for the coordination and the execution.

The direct sales advertisement activities of Sunweb can be divided into push and pull activities. The push activities are used to create demand; Sunweb uses the following push activities: weekly newsletter, segmented mailings to previous customers, special activities like discount coupons, and a brochure. The weekly newsletter and the segmented mailing do not cost extra money. The brochure has some developing and printing costs and price reductions are translated into costs for the marketing department.

For the weekly newsletter, as well as for the daily offer and the highlighted products at the website, Sundio makes use of the system Go Go Marketing, developed in cooperation with ORTEC. The system advises based on the expected increase in profit which products to display on the website and in the newsletter. The expected increase in profit is based on the historical data for the effects of a display and the margin of the products. Besides maximizing the expected increase in profit, some business rules as “only consider last minutes” and “only consider accommodations not shown in last 2 weeks” are taken into account. Sundio makes the final decision on which products to show.

Besides the push activities, Sundio has certain pull activities which help customers who are searching for a product, to find a product at Sunweb. The most important push activity of Sunweb, besides optimizing which products to show as highlighted products at the website, is the use of Google AdWords, affiliate marketing and remarketing.

Google AdWords is a tool of Google that companies can use to bid for advertisements displayed next to the search results of Google. Sundio has to create ads and choose keywords on which they want to advertise. If someone uses a search term in Google related to a keyword chosen by Sundio, Google probably shows the advertisement of Sundio. When a person clicks on an advertisement of Sundio, they are redirected to the related webpage of Sundio.

An example of advertisement positions in Google AdWords is shown in Figure 8. Everytime someone searches for a keyword, all advertisers who selected that keyword are bidding for the different advertisement positions. The position of the advertisement is based on the bid made by the advertiser and the quality score assigned by Google. The quality score is a mark between 0 and 10 and is an indication of how relevant an advertisement, the related webpage and its related keywords are for the person who is shown the advertisement (Google, 2013). Google does not reveal the exact determination method of the quality score.

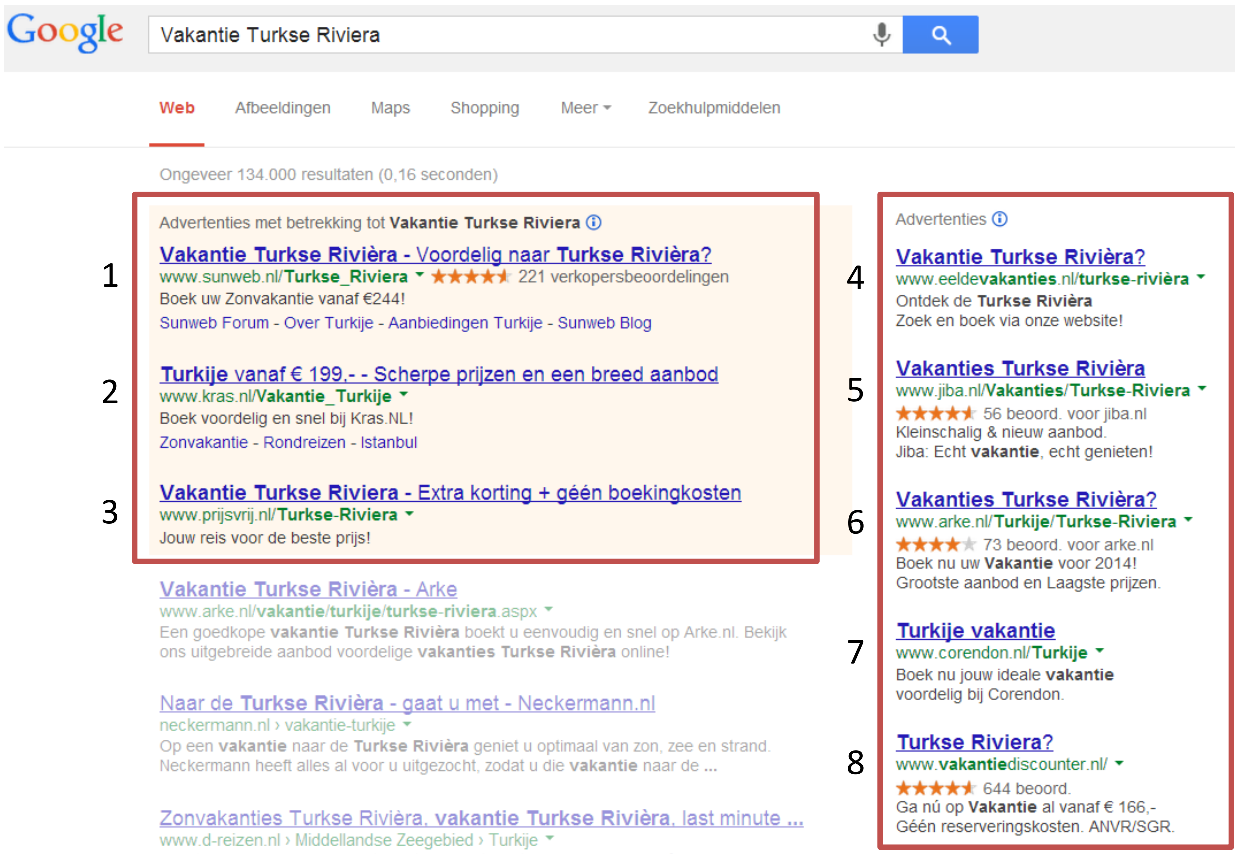

Figure 8: Example of Google AdWords advertisement showing up next to the search results in Google when searching for "Vakantie Turkse Riviera" (in English: "Holiday Turkish Riviera") The numbers indicate the position of the advertisement.

The bid and the quality score of an advertisement are multiplied to get an Ad Rank. The advertisement with the highest Ad Rank is shown at position 1, the second on position 2, and so on. When someone clicks on an advertisement, the related advertiser has to pay the minimum price for which the advertisement has a higher Ad Rank than the advertisment on the next position. We illustrate this in Example 3.3.

Example 3.3. *We give a simplified example based on the data show in Table 2: advertisement A is placed on the first position since it has the highest Ad Rank. When for advertisement A a bid of € 0.90 was made, it would have the same Ad Rank as advertisement B; therefore when someone clicks on advertisement A the related advertiser has to pay € 0.91 for this click.*

Table 2: Example data for explaining the bid system of Google AdWords

Advertisement	Bid (€)	Quality score	Ad Rank	Position
A	1.10	6.0	6.6	1
B	0.60	9.0	5.4	2
C	0.90	5.0	4.5	3

Within Google AdWords Sundio has to draw up a list of keywords on which it wants to advertise. These keywords are organised by grouping them in different ad groups. Different ad groups are at their turn part of a campaign.

Generally Sundio uses five different campaigns for every destination country. These campaigns are related to (i) cities, (ii) regions, (iii) accommodations, (iv) last minute offers or (v) other keywords. These campaigns are made up of multiple ad groups. An ad group is a set of similar keywords and contains one or more ads which can be displayed by Google. For every keyword a maximum cost per click can be set which is the bid for that keyword. Example 3.4 illustrates the structure used in Google AdWords by Sunweb.

Example 3.4. *In this example we consider advertisements for destination country Turkey. For the destination country Turkey the following campaigns are set up:*

- *Accommodations Turkey*
- *Cities Turkey*
- *Regions Turkey*
- *Last minute Turkey*
- *General Turkey*

For example within the campaign "Accommodations Turkey", an ad group is set up for every accommodation that Sunweb offers at their website; within the campaign "Cities Turkey" there is an ad group for every city in Turkey which are a destination of the holidays offered by Sunweb. Within the campaign "Accommodations Turkey" there is for example the ad group "Turkse Rivièra - Antalya - Lara - Hotel Titanic Resort". This ad group is related to the accommodation Hotel Titanic Resort in the city Antalya. This ad group contains the three advertisements as shown in Figure 9.

Figure 9: Advertisements for the ad group "Turkse Rivièra - Antalya - Lara - Hotel Titanic Resort"

This ad group also contains different keywords, for example: "hotel titanic turkey", "titanic resort", and "vakantie hotel titanic resort". If someone uses one of these keywords searching via Google, Sunweb is bidding to show one of the advertisements as shown in Figure 9. When someone clicks on one of these advertisements he is redirected to http://zon.sunweb.nl/turkije/turkse_riviera/antalya-lara/hotel_titanic_resort.htm which is the page for this resort at the website of Sunweb.

GGM is used to optimize the bids set for different keywords in the system of Google AdWords. The height of the bid influences the position of the advertisement: in general a higher bid results in a better position. The position in turn influences the number of clicks to be expected and more clicks lead to a higher number of conversions. Based on the expected number of conversions for different bids and the margin of the different products, the optimal bid for which the expected profit is maximised is calculated.

The AdWords Bid Optimizer Model used by Sundio contains the following variables:

cpc	Cost per click
pos	Weighted position in Google AdWords
imp	Number of impressions for the advertisement
$clicks$	Number of clicks at advertisement
$conv$	Number of conversions from people who clicked at the advertisement
$marg$	Margin per conversion
β_1	Effect of cpc on pos
β_2	Effect of $pos * imp$ on $clicks$
β_3	Effect of $clicks$ on $conv$

In the model the cost per click is used instead of the bid since the historical bids are not saved in the database and cannot be retrieved from Google AdWords. In the model is assumed that a percentage increase in the bid for a key word leads to a similar increase in terms of percentage for the CPC. The number of impressions is defined as the amount of times that the advertisement is displayed. Every week the data of the costs per click, the weighted position in Google AdWords, and the number of impressions, clicks, and conversion is retrieved using the application programming interface of Google AdWords and saved in the data warehouse of Sundio.

The number of conversions is tracked by Google AdWords by tracking how many times someone who clicked on an advertisement booked a holiday. This is based on the last click method, so the AdWords advertisement which was last clicked before some converted is assigned the conversion based on a cookie placed at the customer's computer. It is questionable whether this method gives a good indication of the value of an advertisement since people can delete cookies or make purchases on another computer than used when clicking on an advertisement.

The margin per conversion is based on the average margin in the last few weeks of the bookings for product groups related to the advertisements. To estimate β_1, β_2 and β_3 the model uses the three linear regressions as shown in 3.5-3.7.

$$pos = \alpha_1 + \beta_1 * cpc \tag{3.5}$$

$$clicks = \alpha_2 + \beta_2 * pos * imp \tag{3.6}$$

$$conv = \alpha_3 + \beta_3 * clicks \tag{3.7}$$

The effects are calculated per campaign, ad group or keyword depending on the number of conversions and position fluctuations noticed. The model estimates the position, clicks, and conversion in week 1 (upcoming week) based on data from week 0 (last week) using formula 3.8-3.10:

$$pos^1 = pos^0 + \beta_1 * (cpc^1 - cpc^0) \tag{3.8}$$

$$clicks^1 = clicks^0 + \beta_2 * (pos^1 * imp^1 - pos^0 * imp^0) \tag{3.9}$$

$$conv^1 = conv^0 + \beta_3 * (clicks^1 - clicks^0) \tag{3.10}$$

By writing the formulas 3.10-3.12 as formulas of cpc^1 we are able to optimize the cost per click of the next week. For formula 3.10-3.12 we get the formulas as in 3.11-3.13.

$$pos^1 = A + B * cpc^1 \tag{3.11}$$
$$where\ A = pos^0 - \beta_1 * cpc^0\ and\ B = \beta_1$$

$$clicks^1 = C + D * cpc^1 \tag{3.12}$$
$$where\ C = clicks^0 + \beta_2 * (A * imp^1 - pos^0 * imp^0)\ and\ D = \beta_2 * B * imp^1$$

$$conv^1 = E + F * cpc^1 \tag{3.13}$$
$$where\ E = conv^0 + \beta_3 * (C - clicks^0)\ and\ F = \beta_3 * D$$

Now the cost per click can be optimized for every keyword by maximizing revenues ($= margin * conv^1$) minus costs ($= cpc^1 * clicks^1$).

Maximize:

$$Z = margin * conv^1 - cpc * click^1 \tag{3.14}$$

$$Z = margin * (E + F * cpc^1) - cpc^1 * (C + D * cpc^1) \tag{3.15}$$

$$Z = margin * E + (margin * F - C) * cpc^1 - D * (cpc^1)^2 \tag{3.16}$$

We can optimize Z by setting the derivative over cpc^1 equal to 0. Since the number of impressions for the next week is not known, it is assumed that $imp^1 = imp^0$. This results in the optimal cost per click as in 3.18.

$$\frac{dZ}{dcpc^1} = (margin * F - C) - 2 * D * cpc^1 = 0 \tag{3.17}$$

$$cpc^1 = \frac{1}{2}(cpc^0 + \beta_3 * margin - \frac{clicks^0}{\beta_1 * \beta_2 * imp^0}) \tag{3.18}$$

Eventually restrictions regarding the position or budget can be taken into account. Possible weaknesses in using the above model are that the effects of the decisions of competitors are not taken into account. However it is difficult to implement these since these are not known. Besides it is assumed that the number of impressions in the upcoming week is the same as in the last week. It could be better to use another forecasting method to get a better estimate for the number of impressions in the next week. Also the influence of the CPC on the number of impressions is ignored. Google only shows a limited number of advertisements next to the search results; which advertisements are shown depends on the position assigned to the advertisements. For example when five advertisements are shown, all advertisements with an assigned position higher than fiver are not shown. Therefore, since the position depends on the CPC, also the number of impressions depends on the CPC.

Sundio also uses affiliate marketing; however this is not optimised by the use of GGM. Affiliate marketing is a form of advertisement where a merchant pays affiliates a fee when a customer visits the website of the merchant and is converted in a sale. Sundio makes use of the affiliate networks Tradetracker and Daisycon. These affiliate networks offer their customers the opportunity to promote the holidays of Sundio. If someone clicks on an advertisement about Sundio, a cookie is placed on his computer. Based on these cookies a conversion can be traced back to an affiliate, this affiliate gets a fixed fee for generating this conversion. Optimization of the affiliate marketing expenditures is not part of the scope of the research.

The last direct sales push activity adopted by Sundio is retargeting. Retargeting is a term used for showing advertisement to people who already visited your website. An example is that when someone looks for a holiday to Turkey on the website of Sundio, there is a reasonable chance that on a website like Facebook or Nu.nl he is shown an advertisement of Sundio showing offers for holiday packages with destination Turkey. For remarketing Sundio makes use of the services offered by the display platforms Sociomatic and Criteo. These companies place real time bids on advertisement positions on different website. Sundio has to pay per click on these advertisements. The display platforms are set such that the ads are only shown to people who already visited the website of Sunweb. Optimization of the retargeting expenditures is not part of the scope of the research.

3.4 Opportunities for improvement

In Section 3.1, Section 3.2 and Section 3.3 we discussed the current methods used as support for decision making in the field of pricing and advertising at Sundio. In this section we discuss the opportunities for improvement that we see.

3.4.1 Forecasting

Regarding the method for forecasting the demand there are some possibilities for improvements. Within the step of exponential smoothing the forecasted demand for the same package five weeks before departure has the same weight as he forecasted demand five weeks after departure. Since the forecast for the package with an earlier departure date is probably more accurate, because the number of bookings is closer to the final number of bookings, it could be reasonable to give this forecast a higher weight. Besides this step could be improved by taking the public holidays into account; for example currently when during the autumn holiday 300 bookings are expected in the remaining period, and in the prior week 50 bookings are expected, a large part of the expected remaining bookings for the autumn holiday are shifted to the prior week.

Another opportunity for improvement is related to the prices used within the forecasting method and the determination of the price elasticities. The prices used are based on the revenues regarding the bookings made. Since these bookings can vary in for example travel length, apartment chosen and insurances being included or not, these prices are not a perfect representation for the prices used on the website and therefore seen by the customers.

Regarding the price elasticity calculation the number of bookings, like for example the prices used by competitors, are not taken into account. By comparing the price and the number of PAX with the average price and number of PAX in the surrounding weeks, the price changes are normalised over the booking period. However, it occurs that a change in demand is just related to peaks in the booking seasons, while with calculating the price elasticities these are related to a change in price.

Another possibility for improvement related to the price elasticity is to use different price elasticities for different departure periods. It is reasonable to assume that for different departure periods customers' willingness to pay a higher price is different.

3.4.2 Pricing model

Regarding the pricing model we think that an improvement can be made by using the average price assigned to the remaining booking period by the model. We think that this could lead to an improvement since the ratios used in the model result in an approximation of the optimal price to use in the remaining period. Example 3.8 shows how the expected profit is approximated by using multiple price classes in the model. The same accounts for using multiple CPC classes to approximate the optimal CPC.

Example 3.8. *We consider one product group with variable costs of € 500, a forecast of 1000 visits in the remaining booking period, a forecasted average PAX per booking of 2.5, a reference price of € 680 and a conversion rate related to this reference price of 0.01. The price elasticity of the product group is -2.5.. We use five price classes of respectively € 570, € 635, € 700, € 765, and € 830. For these price classes the model calculates the expected profit. Figure 10 shows the approximated expected profit and the expected profit for our example product.*

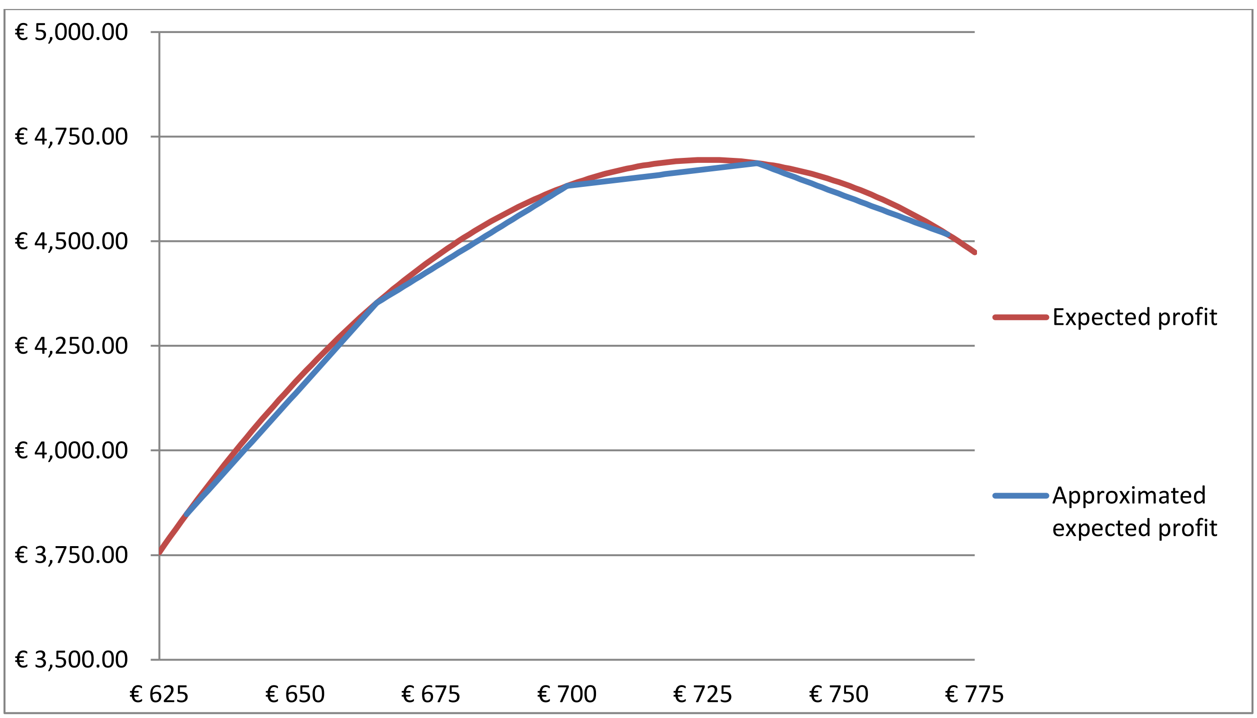

Figure 10: Approximating the expected profit by using multiple price classes

The model assigns a ratio to the different price classes such that the optimal price class is approximated. For our example product the optimal price is € 726, however in this situation the model gives a ratio of 1 to the price class with a price of € 735. When we assume that the capacity available is only 18 PAX, then the optimal price is € 756.16 since a higher price would result in less profit and a lower price would result in a demand which is higher than the capacity and therefore a lower profit. With the capacity restriction of 18 PAX the model would assign a ratio of 0.4 to the price class with a price of € 735 and 0.6 to the price class with a price of € 770. This results in the weighted price of € 756.16 which is exactly the same as the optimal price.

Another improvement related to the pricing model used is to use more time periods. Currently only one time period is used, the prices for the next day are used as input for the model, while it is known that in most cases the price of a package decreases over time. When only one time period is used, the model uses only one price which it expects for all bookings in the remaining booking period. By using multiple time periods it is possible to take into account that the price of a package decreases over time. For the first time period the current price can be used and for the next time periods a price based on the price development of previous years can be used. By doing so the model takes into account that bookings in different time periods are made for different prices. De Vries and Pak (2011) describe the current model used with multiple time periods.

3.4.3 Advertising

When looking at the model used to determine the optimal bid for Google AdWords we also see a possibility for improvement. As mentioned, it is not taken into account that the number of impressions depends on the CPC chosen.

Another disadvantage of the current way of calculating optimal CPCs for the advertisements in Google AdWords is that the capacity is not taken into account. This results in CPCs which are probably not optimal. For example when there are ten keywords related to a product group with a high margin, the CPC advised for these keywords will be high, however when the capacity available is very small for this product group it could be that it is not beneficial to attract many customers via AdWords.

3.4.4 Integrating Revenue Management and Advertising model

Another opportunity for improvement is to integrate both the pricing and the advertisement decisions made. At the moment the prices are based on the current bookings made for a trip, the historical booking patterns and the capacity left. This results in setting lower prices when the demand relative to the capacity left is low or vice versa. By integrating the advertisement decisions it is possible to find opportunities to increase advertisement expenditures and raise prices for a product in order to raise profits. Another result could be that prices for products with a relatively low demand are kept at the same level, while by spending much on advertisement the demand is increased. By integrating advertising decisions and pricing decisions it is possible to find more opportunities for increasing profit.

At the moment the results of the advertising decisions made and the prices set sometimes even counteract each other. The optimization of both the direct sales advertisement activities is based on the margins of the different products. Since the optimization of the advertisements is not integrated into the optimization of the pricing, it is possible that the margin of a product becomes in a downward spiral.

When the demand for a product is low, the price model used advices to decrease the price of the product. Since the margin decreases when the price goes down, the result is that GGM advices to spend less on marketing or to not display this product as a highlight. This results in an even lower demand, and therefore the price of the product decreases even more. When both the effects of changing prices and changing advertisement expenditures are combined in one model, these negative effects of the current system can be solved.

Another reason for integrating the advertisement decisions and the pricing decisions is that when marketing decision are taken into account, another factor influencing the demand can be taken into account. At the moment the marketing decisions are not used in making a forecast for the demand, while it is reasonable to assume that these influence the demand.

3.4.5 This research

Out of the possible improvements mentioned above we think that the largest win can be achieved by integrating the pricing decision making with the decision making for Google AdWords. Therefore within this research we focus on integrating the model for Google AdWords and the model used for making the pricing decisions.

Besides we think that changing the way of choosing the price to use, using the average price assigned by the model instead of the highest price, would be an easy to implement improvement. Therefore we take this improvement into account in Chapter 7 where we perform simulations with different models.

Furthermore we look at the effects of using multiple time periods. We take this method into account when creating an integrated pricing and advertising model and we will perform simulations where we use multiple time periods in the current model used by Sundio as described by De Vries and Pak (2011).

3.5 Conclusions

In this chapter we answered the first sub-question of our research: "What is the current situation regarding pricing and advertising at Sundio?"

In Section 3.1 we described the current forecasting methods used by Sundio. *It is important that the right forecast for the demand in the remaining booking period and the price elasticity is made since this information is used as input for the price optimization model.* The forecast are based on historical data of the last two years. The price optimization model discussed in Section 3.2 indicates which price should be offered for which fraction of the remaining booking period of a product. A LP-model is used to take into account the network of capacity constraints for the different products.

In Section 3.3 we described the current advertisement activities performed by Sundio. Sundio uses different branding activities to create more brand awareness. Besides they use direct sales advertisement to let people find the products offered by Sundio. These can be differentiated into push activities and pull activities. By using GGM, Sundio optimizes these activities in order to generate extra demand for products with a high margin. GGM uses the expected margin for the products. *The most important online advertisement media used by Sundio is Google AdWords. The bids used for the different* *keywords in Google AdWords are optimised based on the expected number of impressions, clicks and conversions of a keyword for different cost per clicks.*

In Section 3.4 we discussed multiple opportunities for improvement. *Integrating both the results of changing prices and advertisement expenditures can result in finding new opportunities for increasing profit.* This is not possible in the current situation since the optimization of advertisement decisions takes place after the optimization of the prices. *The current usage of GGM and GGY can even result in an undesirable downwards spiral.* This can occur when the pricing system advices to lower the price of a product because of the low demand, a result is that less advertisement is performed for the product which even results in a decrease of demand. Besides the model used for calculating the optimal CPC for Google AdWords overestimates the benefits of increasing the CPC for cases where the related product groups have a small capacity.

We see another possibility for improvement by *taking the average price of the prices assigned by the pricing model instead of using the highest price.* Because this improvement is easy to implement we take this improvement into consideration when performing simulations on different models in Chapter 7. *Furthermore we think that using multiple time periods results in a higher profit because the model then takes into account that different prices are used during the remaining booking period*. We also integrate the possibility of using multiple time periods in the Integrated Pricing and Advertising model that we propose.

In the next chapter we answer the second sub-question formulated. We perform a literature review to examine what literature is available related to revenue management and Google AdWords.

4 Literature Review

To develop a new model for the pricing and advertising decisions at Sundio, we performed a literature study. In this chapter we present the results of this study in order to answer our second research question: "What literature is available related to the decision making in advertising and revenue management?" For our literature review we focus on articles which were available for free on the internet or in the collection of the library of the University of Twente.

Since ORTEC already implemented a pricing system at Sundio, we only briefly discuss the existing literature related to Revenue Management in Section 4.1. In Section 4.2 we discuss different forecasting methods which we can use when forecasting values for the parameters used in the new model. We also looked for literature regarding the integration of advertising decision in a revenue management model; however we could find any relevant literature regarding this subject.

4.1 Revenue Management

After several studies about controlled overbooking, Littlewood's study for an airline company about protecting seats for full fare passengers at a level where the expected revenue of the full fare passengers is higher than the price paid for discounted tickets, was the start of the later called revenue management or yield management (McGill & Van Ryzin, 1999). In 1979 the airline schedules and prices were deregulated and, later on, possibilities for using computer systems to control the reservations systems became available. As a results of these major changes airline companies developed OR techniques to regulate problems regarding overbooking, discount allocation, traffic management, modelling passenger preferences and determining yield management performance (Smith, Leimkuhler, & Darrow, 1992).

After the fast development of revenue management in the airline industry, other companies also started implementing revenue management to maximize profit. Talluri and Van Ryzin (2005) describe how revenue management can be practiced in, among others, hotel, tour operator, retailing, broadcasting, electricity generation, and passenger railway industries. Nowadays, a widely used definition of revenue management is "a method which can help a firm sell the right inventory unit to the right type of customer, at the right time, and for the right price" (Kimes, 1989).

Talluri and Van Ryzin (2005) make a distinction between price-based and quantity-based revenue management. Quantity-based revenue management is applied by rationing the availability of products in different segments for which prices are set. Price-based revenue management is applied by raising the price of a product in order to limit the demand. At Sundio the decisions are price based due to the fact that a product only has one price at a time; for every product there is only one class available. Since the price of a package at Sundio can change every day, they are not able to advertise with prices in for example printed advertisements.

Weatherford and Bodily (1992) perceive three conditions for industries to implement revenue management:

- a perishable inventory,
- the possibility of segmenting price-sensitive customers, and
- fixed capacity (not necessary, however a common characteristic).

Sundio has a perishable inventory since they cannot sell a flight or accommodation after the related days have passed. Sundio is able to segment markets since they have multiple brands, destinations, durations for stays, and accommodation classes. Besides they start with offering a high price to customers who want to book their holiday in advance, and they offer lower prices last minute to customers who take the risk that their favourable holiday package is not available anymore. At Sundio the capacity is relatively fixed since they buy seats for flights and rooms in accommodations before they start selling the holiday packages; for some flights and accommodations however it is possible to buy extra capacity later on.

Talluri and Van Ryzin (2005) distinguish four steps in the revenue management process which are performed repeatedly. The first step they mention is the collection of historical data. Based on this data the second step is to estimate parameters of the demand model and to forecast the relevant numbers. The third step is to find optimal values for the decision variables used and finally, the fourth step is to take the optimal numbers into operation. Depending on for example the variability in business conditions or the importance of the output these steps can be performed more or less often. For example companies can choose to start optimizing prices on a weekly basis, and increase the frequency of re-optimization when demand increases nearby the date of service.

The four steps in revenue management processes can also be distinguished in the system of Sundio described in Section 3: (1) historical data is gathered from the data warehouse, (2) a forecast for the demand function is made, (3) optimal price allocations are determined by using a mathematical model, and (4) the optimal prices found are offered to customers via the website. For the new model which we introduce in Chapter 5, we make changes regarding the second and the third step. We split the demand function in an attraction function and a conversion function and the mathematical model has to be adapted to not only optimize the prices, but also the advertisement decisions to be made.

4.2 Forecasting methods

In this research we develop a model which supports the pricing and advertising decisions made by Sundio. To be able to evaluate alternate policies we have to make a forecast for uncertain outcomes in the future. Hopp and Spearman (2008) make a distinction between qualitative forecasting and quantitative forecasting. Qualitative forecasting is based on the expertise of people, while quantitative forecasting methods use mathematical models to make predictions. Since the new model has to evaluate many scenarios of different prices for millions of holiday packages and bids for a large amount of keywords used in Google AdWords daily with a planning horizon which is usually less than one year, we focus on the quantitative forecast method.

Quantitative forecast methods can be categorised in two classes:

- Time series forecasting methods "predict a numerical parameter for which past results are a good indicator of future behaviour, but where a strong cause-and-effect relationship is not available for constructing a causal model" (Hopp & Spearman, 2008).
- Causal forecasting methods "attempt to explain the behaviour of an uncertain future parameter in terms of other, observable or at least more predictable, parameters." (Hopp & Spearman, 2008)

We discuss different time series forecasting methods in Subsection 4.2.1. In Subsection 4.2.2 we pay attention to different causal forecasting methods used in operations research.

4.2.1 Time series forecasting methods

DeLurgio (1998) defines a time series as "a continuous set of observations that are ordered in equally spaced intervals." In this research we look at different time series; for example we investigate the number of visitors at the website of Sunweb and the number of bookings for holidays to Turkey. In order to use a revenue model for decision making we have to make forecast for different time series related to the parameters used. In this subsection we first discuss time series in general, after that we describe different methods for making forecasts for time series.

Different cause-and-effect relationships can influence the pattern of a time series. DeLurgio (1998) describes the following frequently occurring patterns:

- *Random patterns* do not show any systematics or repetition. Sometimes correlated time series can be used to make a forecast for a time series with a random pattern.
- Time series with *trend patterns* show a general increase or decrease over time. When such a trend is expected to continue it should be taken into account when making a forecast. An example of a time series at Sundio which could show a trend pattern is the number of bookings of bookings over the last three years.
- *Seasonal patterns* are a result from recurring periodic events. At Sundio we expect to recognize seasonal patterns in for example the number of bookings or the price of a product over time. Seasonal forecasting models are needed to take the seasonal influences into account.
- Time series with *cyclical patterns* show recurring patterns which are not periodic. These time series are difficult to forecast since the duration of the recurring patterns varies. We do not expect that cyclical patterns show up in the time series of Sundio since the seasonal influences are important.
- In time series with *autocorrelated patterns* there is a strong correlation between the value of a series and the value in previous periods. High-volume influences, trends and seasonality influence the autocorrelation of a time series. In time series with a clear autocorrelation between different values simple methods can be used to forecast one period ahead, however for making a multi period forecast seasonality and trends should be taken into account.

We expect that most time series at Sundio will show trend and seasonal patterns. An example of a time series at Sundio which could show a trend pattern is the number of bookings per week. We expect to recognize seasonal patterns in for example the number of visitors at the website and the conversion rate of a product. Autocorrelated patterns can be seen in time series for the number of visitors of the website or the number of impressions of advertisements (high-volume influences), as well as in time series for the number of bookings or the price of a product (trends and seasonality).

We now introduce different forecasting methods which might be applicable for forecasting values for parameters used in our new model.

Naive methods

Naive methods are the simplest methods for forecasting and assume that recent periods are the best indicators for forecasting the future (Hanke & Reitsch, 1998). The simplest naive method is using the last observation as a forecast for the first future period. This can be written as

$$\hat{Y}_{t+1} = Y_t \tag{4.1}$$

where Y_t is the value of the observation in the last occurring time period and $\hat{Y}_{t+1}$ is the predicted value for the next time period. The model can be adjusted for a trend pattern in the data. Therefore the difference between the last two occurring time periods is added to the forecast:

$$\hat{Y}_{t+1} = Y_t + (Y_t - Y_{t-1}) \tag{4.2}$$

A variation on this is using the rate of change:

$$\hat{Y}_{t+1} = Y_t \frac{y_t}{y_{t+1}} \tag{4.3}$$

It is also possible to take a seasonal pattern into account. For example when considering a time series with time periods for every quarter in a year, it is possible to use the same quarter of last year as a forecast. The seasonal pattern and trend pattern can be combined, for example when considering quarterly data we can use

$$\hat{Y}_{t+1} = Y_{t-3} + (Y_t - Y_{t-4}) \tag{4.4}$$

Here we take into account the difference between this year and last year in the last quarter.

Averaging methods

Averaging methods are similar to naïve methods and assume that differences in past observations are due to random error (Hanke & Reitsch, 1998). The simplest forecast for averaging methods is

$$\hat{Y}_{t+1} = \frac{\sum_{t=1}^{n} Y_t}{n} \tag{4.5}$$

where n is the number of observations already made. It is also possible to take only the last n relevant observations into account. In that case we speak of a moving average which we can write as

$$\hat{Y}_{t+1} = \frac{\sum_{i=1}^{n} Y_{t-i+1}}{n} \tag{4.6}$$

The moving average method can be extended by giving different weights to different observations. Double moving averages can be used in time series with a trend pattern. If we note the moving average of time t as M_t we can use equation (4.7) till (4.10) for calculating a forecast based on the double moving average method.

$$M'_t = \frac{\sum_{i=1}^{n} M_{t-i+1}}{n} \tag{4.7}$$

$$a_t = 2M_t - M'_t \tag{4.8}$$

$$b_t = \frac{2}{n-1}(M_t - M'_t) \tag{4.9}$$

$$\hat{Y}_{t+p} = a_t + b_t * p \tag{4.10}$$

In (4.10) p is the number of periods that the forecast is made in ahead.

Exponential smoothing methods

Exponential smoothing methods take old forecast into account when calculating a new forecast. The smoothing constant is used to give a weight to the previous forecast (Hanke & Reitsch, 1998). The simple exponential smoothing method can be noted as

$$\hat{Y}_{t+1} = \alpha Y_t + (1-\alpha)\hat{Y}_t \tag{4.11}$$

The value of α, used as the smoothing constant, can be chosen by evaluating the forecast method on historical data; α should be chosen such that the error is minimized. We discuss the evaluation of time series forecasting methods later in this subsection.

We can also use the double exponential smoothing method, also known as Brown's method. This method is similar to the double averaging method and can also be used when considering a time series with a linear trend (Hanke & Reitsch, 1998). The equations for Brown's method are formulated in (4.12)-(4.15); the exponentially smoothed value of Y_t is noted as A_t and $\hat{Y}_{t+h}$ indicates the forecasted value at time $t+h$.

$$A'_t = \alpha A_t + (1-\alpha)A'_{t-1} \tag{4.12}$$

$$a_t = 2A_t - A'_t \tag{4.13}$$

$$b_t = \frac{\alpha}{1-\alpha}(A - A'_t) \tag{4.14}$$

$$\hat{Y}_{t+h} = a_t + b_t * h \tag{4.15}$$

Brown's method of double exponential smoothing takes into account the trend of times series. Hanke and Reitsch (1998) describe Holt's two-parameter method which "gives more flexibility in selecting the rates at which the trend and slope are tracked". This double exponential smoothing method is formulated in (4.16)-(4.18)

$$\mu_t = \alpha * Y_t + (1-\alpha)(\mu_{t-1} + b_{t-1}) \tag{4.16}$$

$$b_t = \gamma(\mu_t - \mu_{t-1}) + (1-\gamma)b_{t-1} \tag{4.17}$$

$$\hat{Y}_{t+h} = \mu_t + h * b_t \tag{4.18}$$

In this model b_t is the trend estimate of the time series. The parameter γ is a smoothing constant for trend. Gardner Jr and McKenzie (1985) mention that the trend estimate can be modified; therefore they modify the trend by multiplying b_{t-1} in (4.16) and (4.17) with ϕ. Equation (4.19) changes to

$$\hat{Y}_{t+h} = \mu_t + \sum_{i=1}^{h} \phi^h * b_t \tag{4.19}$$

When $\phi = 0$ no trend is taken into account and when $\phi = 1$ the model is equal to the model of Holt. If ϕ is between 0 and 1, the trend is damped, while when ϕ is larger than 1 the trend is exponential. Using an exponential trend for forecasting multiple weeks in advance is not desirable since all forecasts are made automatically, therefore when this option is considered a constraint should be set on the value of ϕ.

An extensive version of the simple exponential smoothing method is known as the Holt-Winters' method and can be formulated as in equation (4.20)-(4.23) (Yaffee & McGee, 2000). This method takes the trend and the seasonality of a time series into account.

$$\mu_t = \alpha(Y_t - S_{t-p}) + (1-\alpha)(\mu_{t-1} + b_{t-1}) \tag{4.20}$$

$$b_t = \gamma(\mu_t - \mu_{t-1}) + (1-\gamma)b_{t-1} \tag{4.21}$$

$$S_t = \ \delta(Y_t - \mu_t) + (1-\delta)S_{t-p} \tag{4.22}$$

$$\hat{Y}_{t+h} = \mu_t + h * b_t + S_{t-p+h} \tag{4.23}$$

In this model b_t is the trend estimate of the time series and S_t is the seasonality estimate. The length of the seasonality is indicated by p. The parameters γ and δ used in the Holt-Winters' model are smoothing constants for trend and seasonality estimate respectively. Again we can extend this model by using ϕ in the same manner as discussed above.

The model formulated in equation (4.20)-(4.23) is an additive exponential smoothing model, the multiplicative Holt-Winters' model is described in (4.24)-(4.27).

$$\mu_t = \alpha * \frac{Y_t}{S_{t-p}} + (1-\alpha)(\mu_{t-1} + b_{t-1}) \tag{4.24}$$

$$b_t = \gamma(\mu_t - \mu_{t-1}) + (1-\gamma)b_{t-1} \tag{4.25}$$

$$S_t = \lambda \frac{Y_t}{\mu_t} + (1-\lambda)S_{t-p} \tag{4.26}$$

$$\hat{Y}_{t+p} = (\mu_t + b_t t)S_{t-p+h} \tag{4.27}$$

A simple method for choosing the initial values of μ_t, b_t, and S_t based on the first year of data is proposed by Gardner and Dannenbring (1980) (as discussed by Segura and Vercher (2001)). For μ_0 they use the average demand of the first year; b_0 is set to zero and $S_{t-p} = \frac{Y_t}{\bar{Y}_1}$ where $\bar{Y}_1$ is the average of the values for the first year. For the additive model the initial values for the seasonality are calculated as $S_{t-p} = \bar{\bar{Y}}_1 - Y_t$.

Evaluating time series forecasting methods

A common way to evaluate the forecasting performance of different time series forecasting methods is to split the gathered data up into two sets of data. The first part of the data is called the in-sample data and is used to set up the parameters of model. The second part of the data is called the out-sample data and is used for evaluating the effectiveness of a model. When the model is constructed based on the in-sample data forecasts can be made for the data points in the out-sample data and evaluated based on the actual values (DeLurgio, 1998).

Hanke and Reitsch (1998) mention four measures of forecast accuracy; these measures can be used to compare different forecasting methods. The first measure is the mean absolute deviation (MAD) which measures the magnitudes of the forecast errors. The MAD is defined as

$$MAD = \frac{\sum_{t=1}^{n}\left|Y_t - \hat{Y}_t\right|}{n} \tag{4.28}$$

According to Winston (2004) the standard deviation of a forecasting model with normally distributed forecast errors is equal to 1.25 MAD. We know that approximately 95% of the predictions in such a forecasting model should be within the two standard deviations of the actual value.

The MAD is used to express the forecasted errors in the same unit as the time series; the mean absolute percentage error (MAPE) expresses the forecasted errors in terms of a percentages. The MAPE is defined as

$$MAPE = \frac{\sum_{t=1}^{n}\frac{\left|Y_t - \hat{Y}_t\right|}{Y_t}}{n} \tag{4.29}$$

Another method is using the mean squared error (MSE) which squares the forecasting errors. Therefore large forecasting errors have a higher weight. The MSE can be calculated by

$$MSE = \frac{\sum_{t=1}^{n}\left(Y_t - \hat{Y}_t\right)^2}{n} \tag{4.30}$$

4.2.2 Causal/multivariate forecasting methods

In this subsection we discuss different causal forecasting methods. These methods can be used to make a forecast of a parameter based on other observable or more predictable parameters.

Simple regression analysis
A simple regression can be used to model the relationship between two related variables. As described by Winston (2004) a linear relationship between x_i and y_i can be modelled as:

$$y_i = \beta_0 + \beta_1 * x_i + \varepsilon_i \tag{4.31}$$

In this equation ε_i is the error term used representing the difference between $\beta_0 + \beta_1 * x_i$ and y_i. In most cases the true values for β_0 and β_1 are unknown, therefore the values for β_0 and β_1 are chosen such that the sum of squared errors (SSE), calculated as $\sum \varepsilon^2$, is minimized.

When the relationship between x_i and y_i is non-linear, a transformation can be used to generate new predictor variables. The most common transformations are using $\log(x_i)$, x^2, $\sqrt{(x_i)}$ or $1/x_i$ as predictor variables (Hanke & Reitsch, 1998).

A regression analysis procedure is based on three assumptions; the errors should be (i) normally distributed, (ii) independent of each other and (iii) independent of the variable x (Winston, 2004).

For determining the accuracy of predictions based on a simple regression the standard error of the estimate ($\mathrm{s_e}$) can be used. The standard error of the estimate is given by

$$\mathrm{s_e} = \sqrt{\frac{\mathrm{SSE}}{\mathrm{n} - 2}} \tag{4.32}$$

Approximately 95% of the values for y will be within $2\mathrm{s_e}$of predicted value $\hat{\mathrm{y}}$. The observations for which y is not within $2\mathrm{e_s}$ of $\hat{\mathrm{y}}$ can be considered as an outlier (Winston, 2004). The F-test of ANOVA can be used to test the significance of a linear relationship.

Multiple regression analysis
Instead of using one variable for predicting the value of a dependent variable, a multiple regression model can be used. The general form a multiple regression model is (Winston, 2004)

$$y_i = \beta_0 + \beta_1 * x_{1i} + \beta_2 * x_{2i} + \cdots + \beta_k * x_{ki} + \varepsilon_i \tag{4.33}$$

In such a multiple regression model it is possible to introduce dummy variables. Hanke and Reitsch (1998) mention that such variables "are used to determine the relationship between qualitative independent variables and a dependent variable." For example when using a regression model it is possible to introduce a dummy variable which indicates whether the observation was made during a holiday or not.

For a multiple regression model the standard error of the estimate is defined as

$$s_e = \sqrt{\frac{SSE}{n - k - 1}} \tag{4.34}$$

When using a multiple regression model, multicollinearity can be present. Multicollinearity exists when there is a strong relationship between some of the independent variables and may result in unreliable regression coefficients (DeLurgio, 1998). DeLurgio (1998) mentions some potential solutions for multiple regressions where multicollinearity exists: it is important to make sure that no variables have the same underlying causal factor, besides when the multicollinearity is a result of a specific sample it can be useful to add extra observations.

4.3 Conclusions

In this section we discussed the literature related to our research. In section 4.1 we briefly discussed the most important literature in the field of revenue management. Revenue management started in the 1970's in the airline industry. At Sundio price-based revenue management is applied: the capacity for the packages is known and prices have to be set. *Since the capacity of Sundio is perishable and (almost) fixed, and they can segmentate price-sensitive customers by changing the price over time, Sundio is able to apply Revenue Management on their products.*

At Sundio the four steps in the Revenue Management can be distinguished as: (1) historical data is gathered from the data warehouse, (2) a forecast for the demand function is made, (3) optimal price allocations are determined by using a mathematical model, and (4) the optimal prices found are offered to customers via the website. *The new model which we propose in the next section influence step 3, the allocation of optimal prices, since the advertisement decisions are integrated in the model. Therefore also step 2, making a forecast for the demand function, changes since some extra forecast have to be made.*

The different forecasting methods which we discussed in Section 4.2 are used in Chapter 6 when making the new forecasts needed. We make a distinction between time series forecasting methods and causal forecasting methods. The time series methods can be applied when we want to predict parameters where a strong cause-and-effect relationship is not available. We use these methods for example when forecasting the number of visits at the website or the number of impressions in Google AdWords.

The causal models can be used when we want to predict the value of a parameter in terms of another predictable parameter. These methods can be used when for example we want to investigate the relation between the number of impressions for Google AdWords advertisements and the CPC used in Google AdWords. We also explained how the accuracy of the forecasting methods in forecasting the parameters can be measured.

In the next section we discuss how the effects of pricing and advertising decisions can be combined in an integrated model.

5 Integrated Pricing & Advertising Model

In the previous chapter we discussed the literature related to our research. In this chapter we answer the third sub-question formulated in Chapter 2.2: *"How can the effects of the decisions in pricing and advertising be combined in an integrated model?"* Therefore we propose a new model for supporting decision making regarding marketing and pricing at Sundio.

Regarding the advertising decisions we focus on the CPCs set for different AdWords groups. We define an AdWords group as a group of keywords with the same destination URL for the advertisement. Table 3 gives an overview of the structure of destination URLs.

Table 3: Overview and examples of the structure for destination URLs

URL	Example
zon.sunweb.nl	zon.sunweb.nl/
zon.sunweb.nl/[country]	zon.sunweb.nl/turkije/
zon.sunweb.nl/[country]/[region]	zon.sunweb.nl/turkije/turkse_riviera/
zon.sunweb.nl/[country]/[region]/[city]	zon.sunweb.nl/turkije/turkse_riviera/side/
zon.sunweb.nl/[country]/[region]/[city]/[accommodation]	zon.sunweb.nl/turkije/turkse_riviera/side/ appartementen_sun_city.htm

This means that a click on an advertisement for any of the keywords in an AdWords group results in being redirected to the same webpage. We choose to focus on the CPCs set in AdWords since this is the main online advertising platform used by Sundio and there are detailed data regarding the number of impressions, number of clicks, and costs of the advertisements available.

For the CPC in AdWords we use the same assumption as in the model currently used: an increase in terms of percentage for the CPC can be reached by increasing the bid for a keyword with the same percentage. For example, when the model shows that the CPC should be increased with 10%, the bid set in AdWords for the keywords should also be increased with 10%. Besides we assume that the people attracted via AdWords can visit the product page of the product group where they see the price of the product and make a booking or not based on the price of the product group.

As discussed in Section 1.3 the new model should split the demand function in the number of people being attracted to the website and the conversion rate at the website. The number of customers attracted to the website is based on the bids set in Google AdWords, while the conversion rate for customers at the website depends on the price set for a package. We can write the objective function which we want to maximize as

$$profit = \sum_{all\ products} (margin\ per\ pax * \#\ visits\ * conversion\ rate * pax\ per\ booking) - costs\ AdWords \quad (5.1)$$

Compared to the current model used we formulate the demand in PAX as the number of visitors multiplied with the conversion rate and the number of PAX per booking. By taking the sum over all products over the number of PAX booked times the margin per PAX the profit on bookings is calculated. Note that this margin does not include the costs for Google AdWords; these costs are subtracted from the profit on bookings in order to calculate total profit. This objective function can be used to maximize the result of the price and advertisement decision in terms of profit.

The margin and the conversion rate used in the objective function depend on the prices set for the different product groups; the number of visits and the costs for AdWords depend on the CPC set for the different AdWords groups. An important constraint in the model is the capacity available for the different resources.

In this chapter we start with a model for a very simplified case and we improve this model in order to come to a model which can be used by Sundio. In Section 5.1 we discuss a continuous model for a case where we consider only one product group, one AdWords group, no capacity restriction and one price and CPC to set for the remaining period. The next model, described in Section 5.2, is a linear programming model taking into account the capacity restriction. In Section 5.3 we introduce a model taking into account that different prices and different CPCs can be used over time in the remaining booking period. We improve the model to take into account multiple product and AdWords groups in Section 5.4. We conclude this chapter by summarizing the results and drawing a conclusion in Section 5.5.

For the different models discussed in this chapter we introduce multiple parameters which have to be forecasted. For parameters which are not used in the current model we explain in Chapter 6 how these parameters can be forecasted. Besides we make multiple assumptions:

- Visitors coming from AdWords have the same conversion rate as visitors not attracted via AdWords. This assumption is made for simplicity and because of the fact that there is no reliable data available about whether a booking is made by a visitor attracted via AdWords or not.
- The average number of PAX per booking is not affected by the price used. One could imagine that when someone wants to book a holiday for more people, he is more price aware than when he makes a booking for one person; however for simplicity we do not take this effect into account.
- The effect on the click through rate of changing the CPC is constant over time. Note that since the number of impressions for an AdWords groups changes over time, the number of visits from AdWords as a results of the CPC is not constant over time.
- We assume that visitors from AdWords only visit webpages which are at a lower level than the destination URL of the advertisement they clicked on. For example when someone clicks on an advertisement with the destination URL zon.sunweb.nl/turkije/turkse_riviera/ we do not take into account that he can also visit zon.sunweb.nl/turkije/egeische_kust/ during his visit. We think this assumption is reasonable since the destination URL of an advertisement is related to the search query of the visitors and therefore is primarily interested in holiday offers related to that search query.

5.1 One product group and one AdWords group

In this section we discuss a model where only one product group and one AdWords group are considered. In this model we consider only one price and CPC to be used in the remaining booking period; besides we do not take capacity restrictions into account. The objective of the model is to set a price and CPC which results in the highest expected profit for the remaining booking period.

5.1.1 Notation

In this subsection we discuss the notation used in describing the model in Subsection 5.1.2. We first introduce the following parameters:

Parameters:

q	Sum of variable flight and accommodation costs
rp	Reference price of the remaining period
cr	Forecasted reference conversion rate when using the reference price
e	Price elasticity
b	Baseline visitors
α	Intercept component for the number of visitors attracted via AdWords
β	Effect of the CPC on the number of visitors attracted via AdWords
pb	Forecasted average PAX per booking for the product group

As in the model currently used we only consider the variable costs of a product. We also use a reference price which is the price expected based on the price changes in previous years as explained in Appendix A. For this reference price we use the forecasted conversion rate. The price elasticity indicates how much the conversion rate increases in terms of percentage for every percent that the price used is higher than the reference price. Example 5.1 shows how the price elasticity can be used to calculate the expected conversion rate when using a certain price.

Example 5.1. *We consider a conversion rate for our product group of 0.008, a price elasticity of -2.5 and a reference price of € 600. This means that when we use a price of € 600 0.8% of the people visiting the webpage related to the product group make a purchase. When we increase the price with 5% to a price of € 630, the conversion rate expected decreases with 5%*2.5=12.5%, so when using a price of € 630 we expect a conversion rate of 0.008*0.875=0.007. We can calculate this as*

$$exp\ conv\ rate = cr(1 + e * \frac{new\ price - rp}{rp}) = 0.008(1 + \frac{630 - 600}{600} * -2.5) = 0.007 \tag{5.2}$$

The number of baseline visitors is the number of visitors that we expect when not taking into account the visits from AdWords. The parameters α and β are based on the following linear regression:

$$visits\ from\ AdWords = \alpha + \beta * CPC\ used \tag{5.3}$$

Note that α is expected to be zero since when we do not pay for visitors from AdWords we do not expect to get any visitors from AdWords. However we use the linear regression for a local estimation and therefore when applying the linear regression on a small range of CPCs it is possible that α is not zero. The forecasted average PAX per booking is based on the average PAX per booking in previous years.

We use the following notation for the decision variables.

Decision variables:

P	Selling price of the product group
CPC	CPC set in AdWords for the remaining booking period

The variable P indicates the new selling price of the product group. The variable CPC is used as the new average CPC for the AdWords group. Both variables have a range of [0,∞).

5.1.2 Model

Now that we have declared the different parameters and variables we express the objective function, mentioned in (5.1), in terms of these parameters and variables. We use

$$margin\ per\ pax = P - q \tag{5.4}$$

The margin is simply the price set minus the costs. The number of visits can be calculated by taking the number of baseline visits and adding the number of visits from AdWords:

$$\#visits = b + \alpha + \beta * CPC \tag{5.5}$$

We calculate the expected conversion rate based on the price elasticity, the difference between the price used and the reference price in terms of percentage and the reference conversion rate:

$$conversion\ rate = cr * (1 + e * \frac{P - rp}{rp}) \tag{5.6}$$

The expected costs for AdWords can be calculated as the cost per clicks times the number of clicks expected for the CPC.

$$costs\ AdWords = CPC * (\alpha + \beta * CPC) \tag{5.7}$$

The costs for AdWords are the CPC used times the number of visits from AdWords. By using the equations from (5.4)-(5.7) and the parameter for the average number of PAX per booking to formulate the following objective function:

$$\max_{P,CPC} (P - q) * (b + \alpha + \beta * CPC) * \left(cr * \left(1 + e * \frac{P - rp}{rp}\right)\right) * pb - CPC * (\alpha + \beta * CPC) \tag{5.8}$$

This objective function is only limited to the constraint that the CPC and the price set should be higher than 0. In Section 5.2 we extend the model with a capacity restriction. In the next subsection we discuss how the model can be solved and the usability of the model.

5.1.3 Discussion

In this subsection we discuss the optimal solution for the model proposed in the previous subsection and we discuss the applicability of the model.

To find the optimal solution for the model proposed in the previous subsection we first calculate the partial derivative of the objective function with respect to the price. We calculate the optimal price by setting the derivative of the objective function over the price equal to zero. This results in

$$P^* = \frac{1}{2}(rp + q - \frac{rp}{e}) \tag{5.9}$$

where P^* is the optimal price for the model (step by step solution in Appendix C). Since the price elasticity is assumed to be negative the optimal price would not become negative. We want to remark that the optimal price is independent of the CPC set; note that this changes when we take the capacity constraint into account. Now we can also calculate the optimal value for the CPC by filling in the optimal price in the objective function and setting the derivative over the CPC equal to zero. We find (step by step solution in Appendix D) that the optimal CPC, noted as CPC^*, can be described as

$$CPC^* = \frac{1}{2}(M - \frac{\alpha}{\beta}) \quad (5.10$$

where

$$M = (P^* - q) * (cr * \left(1 + e * \frac{(P^* - rp)}{rp}\right) * pb \tag{5.11}$$

We can describe M as the margin per visit when using the optimal price. We see that the optimal CPC depends on the price chosen. This is also what we expected since the price it is worth to pay for a visitor depends on the profit you expect from getting this visitor. In Example 5.1 we give an example for the optimal price and CPC for a certain case.

Example 5.1. *In this example we consider a product group with a reference price of € 700, a conversion rate at the reference price of 0.002, a price elasticity of -3.5, variable costs of € 400, 1000 baseline visits expected in the remaining booking period, and an average number of PAX per booking of 2.6. The AdWords group considered has 5000 impressions, $\alpha = 0.02$, and $\beta = 0.075$. Figure 11 shows a heat map of the profit for using different combinations of prices and CPCs.*

Taking no capacity constraint into account the optimal price is calculated as

$$P^* = \frac{1}{2} * \left(€\,700 + €\,400 - \frac{€\,700}{-3.5}\right) = €\,650 \tag{5.12}$$

With this optimal price we can calculate the expected margin per visitor as

$$M = (€\,650 - €\,400) * (0.002 * \left(1 - 3.5 * \frac{(€\,650 - €\,700)}{€\,700}\right) * 2.6 = 1.65 \tag{5.13}$$

Now that we know the expected margin per visitor we calculate the optimal CPC as

$$CPC^* = \frac{1}{2}\left(€\,1.65 - \frac{0.02}{0.075}\right) = €\,0.69 \tag{5.14}$$

In the right plot of Figure 11 we see how the optimal CPC depends on the price of the product group.

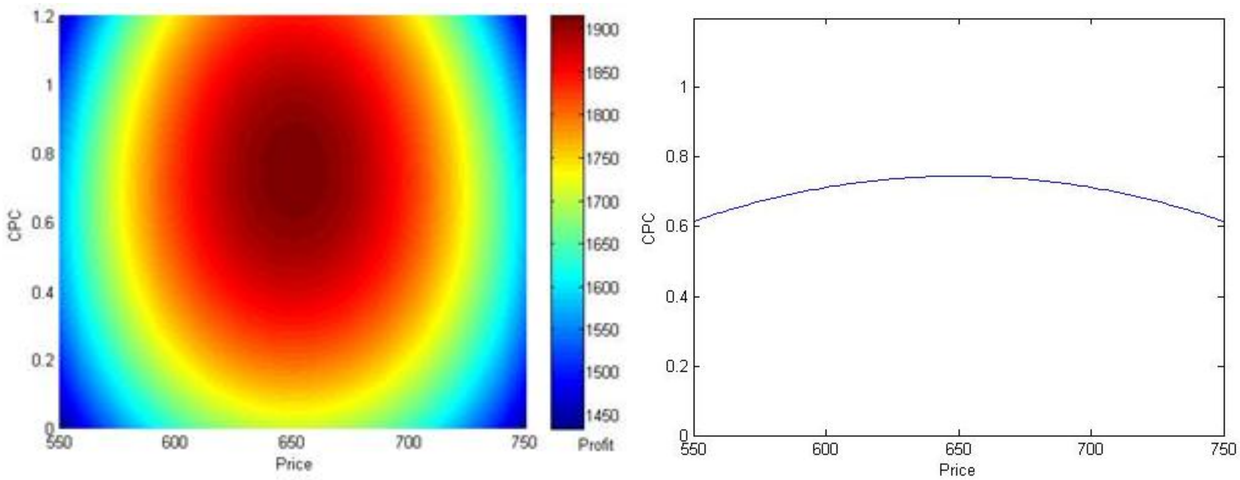

Figure 11: Heat map showing the profit for different prices and CPCs (left) and a plot of the optimal CPC for different prices (right)

The model described gives an optimal solution for a situation in which no capacity constraint and only one AdWords and product group are considered. Since at Sundio the number of PAX is limited to multiple capacity constraints (per flight and per night at an accommodation) this model is not applicable to optimize the prizes of all product groups at Sundio. The optimal price and optimal CPC change when the capacity constraint is taken into account and therefore this model is only applicable for cases where the expected demand is much lower than the lowest capacity constraint of product group considered. The calculation of the optimal CPC can be used when considering only one AdWords group and product groups for which the capacity forms no constraint; however the calculation of M changes since the expected margin should be calculated over multiple products. In the next subsection we extend the model by taking the capacity constraint into account.

5.2 Adding the capacity constraint

In the previous section we introduced a continuous model considering no capacity constraint, one product group and one AdWords group. In this section we add a capacity constraint to the model. Since in this chapter we are working towards a model which can take multiple product groups and AdWords groups into account and therefore has to optimize many thousands of variables, we already transform our model to a linear programming model in this section. The model described in this section is quite similar to the current model as described in Section 3.2.

5.2.1 Notation

In this section we introduce the notation used in describing the model in Section 5.2.2. First we introduce the indices, then we mention the parameters and we end with describing the variables used.

Indices:

j Index of price classes

h Index of CPC classes

Since we only consider one product group and one AdWords group in this section, we do not need indices for the product groups and AdWords groups. Compared to the model currently used by Sundio, as described in Section 3.2, we added an index for the CPC classes. In the model we use multiple CPC classes indicating which CPC to use for an AdWords group. An example of CPC classes could be a change in the CPC with 0%, 5%, 10%, resulting in five CPC classes since a change can result in a higher or lower CPC. Besides a CPC class should be added with zero costs, indicating that no people have to be attracted via AdWords. This is needed in case that a price class is assigned but no visits from AdWords have to be attracted.

Parameters:

p_j Selling price per PAX of the product group when using price class j

q Sum of variable flight and accommodation costs per PAX

cpc_h Average CPC for the AdWords group in CPC class h

b Forecasted number of baseline visits in the remaining booking period

a_h Forecasted number of visits from AdWords when using CPC class h in the remaining booking period

cr_j Forecasted conversion rate when using price class j

pb Forecasted average PAX per booking for the product group

c Units of capacity available for the bottleneck resource

The first two parameters are also used in the current model, however now the index indicating the product group is not needed. We introduce the parameters cpc_h indicating what the average CPC in the AdWords group will be when using price class h.

The parameter b indicates the forecasted number of baseline visits for the product group. The parameter a_h indicates how many extra visits are attracted when using CPC class h for the AdWords group. We introduce the parameter cr_j which indicates the conversion rate when using price j. The parameter pb indicates the forecasted average PAX per booking for the product group.

As in the current model we use a parameter for the amount of units of capacity available. Since we only consider one product group in this section, we only need to know the capacity of the resource with the lowest capacity available.

Decision variables:

$X_{j,h}$ Fraction of the remaining booking period that the combination of price class j and CPC class h is used

The variable $X_{j,h}$ is similar to the variable used in the current pricing model; we add the index h so that we can determine the fraction of the remaining booking period that a combination of a certain price and CPC is used. In the current situation the highest price for which a ratio is assigned is used. We propose to use a weighted average of the prices assigned to the remaining period. The reason for this is that using different price classes is only a way of approximating the optimal value as explained in Section 3.4. Since the optimal CPC is also approximated by using multiple classes for different CPCs we also use a weighted value for the CPC. An example of how to interpret the outcomes of the model is shown in Example 5.2.

Example 5.2. *For this example we use that* $p_3 = 700$ *and* $p_4 = 735$*, besides we have* $cpc_2 = 0.76$ *and* $cpc_3 = 0.8$*. We assume that the outcome of the model is* $X_{3,2} = 0.45, X_{3,3} = 0.15$ *and* $X_{4,3} = 0.4$*. This means that according to the outcomes of the model we should use a price of € 700 for a fraction of 0.45+0.15=0.6 of the remaining booking period. Besides we should use a price of € 735 for 40% of the remaining period. The price that we advise to used based on these outcome is 0.6*€ 700+0.4*€ 735=€ 714. Besides the CPC class two and three should be used for 45% and 55% of the remaining booking period respectively. The weighted CPC would be 0.45*€ 0.76+0.55*€ 0.8=€ 0.782.*

Note that the results of the model improve when more price classes are used; however this also increases the calculation time. Another advantage of using a weighted price and CPC is that we do not have to think about which price and which CPC to use first. Since the optimal price is recalculated every day taking into account the reference price, which in most cases is declining over time, the price is also decline over time.

5.2.2 Model

Now that we have declared the indices, parameters and variables used we can formulate the objective function (see equation (5.1)) in terms of these parameters and variables. We can calculate the margin in the same way as how it is calculated in the current model, simply by subtracting the costs from the price used:

$$margin_j = (p_j - q) \tag{5.15}$$

For the objective function we also need to calculate the number of visits. The number of visits depends on the CPC class used and can be calculated as

$$\#\, visits_h = b + a_h \tag{5.16}$$

The number of visitors is calculated by taking the sum of the number of baseline visitors and the number of people attracted via AdWords.

The costs for Google AdWords can be calculated by multiplying the number of visitors attracted via AdWords with the CPC paid for those visitors. This is multiplied with whether a CPC class is used for the AdWords group:

$$costs\ AdWords = \sum_h \left(a_h * cpc_h * \sum_j X_{j,h} \right) \tag{5.17}$$

By filling in the formulas from equation (5.15)-(5.17) in the objective function as formulated in equation (5.1) and using the parameters cr_j and pb_t we come to the following objective function for our model:

$$Max \sum_j \sum_h \left(\left(p_j - q\right) * (b + a_h) * cr_j * pb * X_{j,h} \right) - \sum_h \left(a_h * cpc_h * \sum_j X_{j,h} \right) \tag{5.18}$$

Note that we can simplify the objective function by using

$$profitbookings_{j,h} = \left(p_j - q\right) * (b + a_h) * cr_j * pb \tag{5.19}$$

and

$$adcosts_h = a_h * cpc_h \tag{5.20}$$

With the parameter $profit_{j,h}$ and $adcosts_h$, which only depend on other parameters, we can write the objective function as

$$Max \sum_j \sum_h \left(profitbookings_{j,h} - adcosts_h \right) * X_{j,h} \tag{5.21}$$

In the objective function the expected profit for using a combination of a certain price and CPC is calculated. This is multiplied with the fraction that this combination is used. The objective function is subject to multiple constraints. The first constraint is a capacity constraint and similar to the capacity constraint used in the current model used:

$$\sum_j \sum_h (b + a_h) * cr_j * pb * X_{j,h} \leq c \tag{5.22}$$

We calculate the total demand for the product group by taking the sum over price classes and CPC classes of the number of people attracted multiplied with the conversion rate using a certain price and the number of PAX per booking. This results in the total expected number of PAX to be booked, which has to be smaller or equal than the capacity available (since Sundio does not allow overbooking).

As in the current model used we have to implement a constraint to make sure that the sum of the fractions is smaller or equal to one and that each fraction is not smaller than zero.

$$\sum_j \sum_h X_{j,h} \leq 1 \tag{5.23}$$

$$X_{j,h} \geq 0 \qquad \forall\, j, h \tag{5.24}$$

By allowing the number of price and CPC classes used in a time period to be smaller than one it is possible to assign no price to a product. This would only be the case if the margin for every price class is negative or if the capacity is reached. By not forcing the sum of the fractions to be exactly one the model is also feasible in situations when using the highest price class for the remaining booking period would result in a demand higher than the capacity.

5.2.3 Discussion

In this section we discuss the application of model proposed in the previous subsection. We have implemented the model in AIMMS 3.13; Appendix E shows the text representation of the model in AIMMS. Using CPLEX 12.6, AIMMS is able to solve a problem with five price classes and five CPC classes within a hundredth of a second.

There are multiple disadvantages for the model proposed in Section 5.2.2. The model is based on the assumption that the number of visits via AdWords is constant over time; however the number of people attracted with a specific CPC depends on the number of impressions, which is not uniformly distributed over time. Therefore it would be interesting to add a time element to the model discussed. The same accounts for the conversion rate, the number of baseline visits, the reference price and the average number of PAX per booking; all these parameters change over time. Therefore it is interesting to add a time element in order to model these changes over time.

Another important disadvantage is that network effects, as discussed in Section 3.2 are not taken into account; not taking into account the fact that multiple product groups compete for the same capacity results in a sub optimal solution. To be able to use this model the capacity available should be assigned to the different product groups using a resource in advance. Since there are many product groups which all make use of multiple capacities, this would probably not be optimal and take a lot of time. However, the model is still not applicable for Sundio in that case, since the effect of the AdWords group is not correctly taken into account. Visitors coming from AdWords can choose between products from multiple product groups, however in this model the expected margin resulting from the other product groups is not taken into account. For example when a visitor comes at the webpage after clicking on an advertisement for the Hotel Titanic Resort, the can choose between packages with different departure times and different durations. When only the margin from one product group is taken into account, the effects of an AdWords advertisement are underestimated.

Another disadvantage is that only one AdWords group is taken into account. In reality a product group is affected by multiple AdWords groups, besides most AdWords groups affect the number of visitors for multiple product groups. Since a product group is affected by multiple AdWords groups it is possible to attract the visits via the most profitable AdWords groups (lowest costs per expected conversion); since we only take one AdWords group into account, the model is not able to choose between multiple AdWords groups to attract visitors.

Since the model does not take into account that the number of visitors for multiple product groups can be affected by an AdWords group, the effect of raising the CPC is underestimated. For example when the CPC for an AdWords group with destination page zon.sunweb.nl/turkije/ is raised we expect that the number of visitors, and therefore the number of bookings, of multiple product groups will increase.

In the next section we add the possibility of considering multiple time periods to the model.

5.3 Using different prices and CPCs for multiple time periods

In the previous section we improved our model by taking into account the capacity restrictions. In this section we extend the model by adding the possibility to consider multiple time periods. This makes it possible to adjust the price and the CPC based on the number of visitors and impressions in AdWords expected. The model currently used by Sundio considers only one time period; De Vries and Pak (2011) describe the same model taking into account multiple time periods.

5.3.1 Notation

In this section we discuss the notation used for describing the extended model in the next subsection. In this section we only describe the indices, parameters and variables that changed compared to the model described in the previous section. We start with introducing an index for time periods

Indices:

t Index of time periods

The index t for time periods is not used in the current model; however De Vries and Pak (2011) already propose using an index for time periods in their pricing model for an online tour operator. The time periods are partitions of the remaining booking period of the product. Note that the time periods do not have to be of the same length. For example it is possible to use a time period for the upcoming week and one for the rest of the remaining booking period. In this case the model proposes an optimal price and CPC for the next week (instead of for the whole remaining booking period as does the model in the previous section) taking into account the rest of the booking period, while the number of variables and parameters is limited.

Parameters:

$p_{t,j}$ Selling price of the product group when using price class j in time period t
b_t Forecasted number of baseline visits for the product group in time period t
$a_{t,h}$ Forecasted number of visits from AdWords when using CPC class h in time period t
$cr_{t,j}$ Forecasted conversion rate when using price class j in time period t
pb_t Forecasted average PAX per booking for the product group in time period t
c Units of capacity available

The parameters p_j, q, cpc_h, and c are identical to the parameters discussed for the model without a time element. For the other parameters we added the index t so that these can be defined per time period.

Variables:

$X_{t,j,h}$ Fraction of time period t that the combination of price class j and CPC class h is used

The variable used in this model is similar to the variable used in the model described in the previous subsection; we only add the index t such that for each time period a price and CPC can be chosen. Again we propose to use the weighted price and CPC as the price and CPC to use in the next period.

5.3.2 Model

Now that we have declared the indices, parameters and variables used we can formulate the objective function in terms of these parameters and variables. The objective is similar to the objective function for the model without a time model, except that we now have to take the sum over all time periods.

$$Max \sum_t \left(\sum_j \sum_h \left(\left(p_{t,j} - q\right) * \left(b_t + a_{t,h}\right) * cr_{t,j} * pb_t * X_{t,j,h} \right) - \sum_h \left(a_{t,h} * cpc_h * \sum_j X_{t,j,h} \right) \right) \tag{5.25}$$

Note that we can simplify the objective function again, therefore we use

$$profitbookings_{t,j,h} = \left(p_{t,j} - q\right) * \left(b_t + a_{t,h}\right) * cr_{t,j} * pb_t \tag{5.26}$$

and

$$adcosts_{t,h} = a_{t,h} * cpc_h \tag{5.27}$$

With the parameter $profit_{t,j,h}$ and $adcosts_{t,h}$ (which only depend on other parameters) we can write the objective function as

$$Max \sum_t \sum_j \sum_h \left(profitbookings_{t,j,h} - adcosts_{t,h}\right) * X_{t,j,h} \tag{5.28}$$

This objective function is subject to multiple constraints. The first constraint is a capacity constraint and similar to the capacity constraint used in the model without a time element:

$$\sum_t \sum_j \sum_h \left(b_t + a_{t,h}\right) * cr_{t,j} * pb_t * X_{t,j,h} \leq c \tag{5.29}$$

We calculate the total expected number of PAX for the product group by taking the sum over all time periods, price classes and CPC classes of the number of people attracted multiplied with the conversion rate using a certain price and the number of PAX per booking. The total demand should be smaller or equal than the capacity available. Again need a constraint to ensure that the sum of the fractions assigned to price classes and CPC classes is not higher than one.

$$\sum_j \sum_h X_{t,j,h} \leq 1 \qquad \forall t \tag{5.30}$$

As for the model in the previous section we also need a constraint to ensure that the fractions are non-negative.

$$X_{t,j,h} \geq 0 \qquad \forall t \tag{5.31}$$

5.3.3 Discussion

In this section we discuss the application of model proposed in Section 5.3.2 for Sundio. We have implemented the model in AIMMS 3.13; Appendix F shows the text representation of the model in AIMMS. Using CPLEX 12.6.

A big improvement compared to the model without a time element is that we can use different conversion rates and the average number of PAX per bookings over time for the remaining booking period. The same accounts for the number of impressions for the advertisements in AdWords. With the model with a time element it is possible to take the fluctuation of these parameters into account and optimize the decisions variables based on these fluctuations. For example it can be more profitable to use a higher price or attract more people when the conversion rate is high.

Another advantage of considering multiple time periods is that the model esteems the price changes in the future. In the previous model considers the same price per PAX for a booking close to the departure time as for a booking made early in advance. Since (in most cases) the reference price is declining over time, bookings made later in the booking period have a lower margin. The model proposed in this section takes this effect into account.

A disadvantage of the model proposed is, that we need more specific information about the different parameters. We expect that when for example we need to forecast the number of visits per week instead of over the remaining period, the accuracy of the forecast decreases. Another disadvantage is that we only consider one product and one AdWords group. We should take the network effects as discussed in Section 3.2 into account in order to optimize the expected profit over all products. When these network effects are not taken into account, the capacity constraints are not realistic. As also discussed for the previous model we should also take into account most AdWords groups affect the number of visits for multiple product groups. Not taking this into account results in the effects of AdWords being underestimated.

In the next section we introduce a model similar to the model discussed in this section, but for multiple product and AdWords groups.

5.4 Multiple product and AdWords groups with time element

Because we only considered one product group in the previous three sections, we did not have to take the network of resources into account. In this section we improve the model from the previous section by considering multiple product groups and multiple AdWords groups. An AdWords group can have effect on multiple product groups and the number of visitors from one product group can be affected by multiple AdWords group. For example the number of visits for a product group with holidays to Side in the Turkish Riviera can be affected by the AdWords group with all advertisements with a destination URL zon.sunweb.nl/turkije, but also from advertisements with zon.sunweb.nl/turkije/turkse_riviera or zon.sunweb.nl/turkije/turkse_riviera/side as their destination URL. On the other hand advertisements leading to zon.sunweb.nl/turkije affect the number of visitors of all product groups related to Turkey. Therefore the model proposed in this section is more complex than the models proposed in the previous sections.

5.4.1 Notation

In this subsection we discuss the notation used for the model proposed for optimizing multiple product and AdWords groups. We only discuss the indices, parameters and variables which are different from the ones previously used in this chapter.

<u>Indices:</u>

i Index of product groups

g Index of AdWords groups
k Index of resources

Again we use indices for the different time periods, price classes and CPC classes. We introduce an index for the different product groups which is identical to the index for product groups in the current model used. We introduce the index g for different AdWords groups. As explained in the introduction of this chapter we define AdWords groups based on the destination URL of the advertisements: different AdWords groups with the same destination URL are in the same AdWords group. The index for resources is also used in the model currently used by Sundio. To limit the number of AdWords group we take all keywords related to accommodations from the same product group as one AdWords group.

<u>Parameters:</u>

$p_{t,i,j}$ Selling price of product group i when using price class j in time period t
q_i Sum of variable flight and accommodation costs for product group i
$cpc_{g,h}$ Average CPC for AdWords group g in CPC class h
$b_{i,t}$ Forecasted number of baseline visits for product group i in time period t
$a_{t,g,h}$ Forecasted number of visits from AdWords when using CPC class h for AdWords group g in time period t
$adr_{t,i,g}$ Forecasted ratio of people attracted by AdWords group g that visit a webpage related to product group i in time period t
$rd_{t,i}$ Relative duration that product group i is available for sale in time period t
$cr_{t,i,j}$ Forecasted conversion rate when using price class j for product group i in time period t
$pb_{t,i}$ Average PAX per booking for product group i in time period t
c_k Units of capacity available for resource k

Most parameters used are similar to the parameters used in the previous section, with the addition of an index for product groups or AdWords groups. Besides we introduce the parameter $adr_{t,i,g}$ which indicates what ratio of the visitors attracted via AdWords group g is expected to visit a product page related to product group i. This variable is needed because people attracted by an advertisement in AdWords do not visit all related product groups. For example someone who is attracted by an advertisement with the destination URL at country level will not visit all product groups related to the country, but only a few of them. In the previous models this ratio was already taken into account for the forecasted number of visits from AdWords.

We introduce the parameter $rd_{t,i}$ which indicates the relative duration that a product group is available for sale in a time period. This parameter has the value 1 for all time periods before the time period in which departure takes place and is zero for all time periods after the time period in which departure takes place. For the time period in which departure takes place the relative duration is equal to the fraction that the product is available for sale in this time period. For example when a time period considers the period from the 1st of July till the 8th of July, and the departure date of a product group is the 5th of July, this product group has a relative duration of 4/7 since the product is only available 4 out of 7 days in this time period. In the time periods before the 1st of July the relative duration is 1. The relative durations . In the model for multiple product groups and AdWords groups we also use a set of product groups using resource k, which is also used in the current model of Sundio. Besides we introduce a set of product groups related to AdWords group g.

Sets:

S_k Set of product groups using resource k

G_i Set of AdWords groups related to product group i

The set S_k consists of every product group related to resource k. As in the current model used a resource can be a night at a specific accommodation or a flight to or from the destination. The set G_g consists of all AdWords groups related to product group i. For example the AdWords group with the destination URL leading to the web page of that country is related to all product groups with the same destination country, while the AdWords group with the destination URL of the Turkish Riviera is only related to product groups with a destination in the Turkish Riviera.

Variables:

$X_{t,i,j,g,h}$ Fraction of time period t that the combination of price class j and CPC class h are used for product group i and AdWords group g

$Y_{t,i,j}$ Fraction of time that product i is used in time period t that price class j is used for product group i

$Z_{t,g,h}$ Fraction of time period t that CPC class h is used AdWords group g

We use a similar variable as in the previous section; the only change is that we add the indices for the AdWords groups and the product groups. Besides we introduce a variable which indicates whether a price class is used for a product group and a variable indicating whether a CPC class is used for an AdWords group. These two variables are needed to not double count visitors from AdWords or costs. The variable are related such that $X_{t,i,j,g,h} = Y_{t,i,j} * rd_{t,i} * Z_{t,g,h}$.

5.4.2 Model

Now that we have declared the indices, parameters and variables used we can formulate the objective function in terms of these parameters and variables. Since the model is much more complex than the previous models described, we again first show how the different parts of the objective function in equation (5.1) are modelled. Again, the margin is just the price minus the costs:

$$margin_{t,i,j} = (p_{t,i,j} - q_i) \tag{5.32}$$

For the objective function we also need to calculate the number of visits. The number of visits depends on the CPC class used and can be calculated as

$$\#\, visits_{t,i,j} = \left(b_{i,t} * Y_{t,i,j} + \sum_{g \in G_i} \sum_{h} a_{t,g,h} * adr_{t,i,g} * X_{t,i,j,g,h} * rd_{t,i} \right) \tag{5.33}$$

The number of visitors is calculated by taking the sum of the number of baseline visits times the ratio that the price is used plus the number of people attracted via AdWords. The number of people attracted via AdWords is the sum over all AdWords groups related to the product and the different CPC classes of the number of clicks times the ad ratio times the ratio that the CPC class is used in combination with a certain price. This means that per AdWords group the model calculates how many visits are expected; this is multiplied with the ratio of people that will visit a web page of the product group. The sum of the baseline visits and the number of visits from AdWords is the total number of visits for the product group using price class j.

The costs for Google AdWords can be calculated by multiplying the number of visitors attracted via AdWords with the CPC paid for those visitors. This is multiplied with whether a CPC class is used for the AdWords group:

$$costs\ AdWords = \sum_t \sum_g \sum_h a_{t,g,h} * cpc_{g,h} * Z_{t,g,h} \tag{5.34}$$

By filling in the formulas from equation (5.28)-(5.30) in the objective function as formulated in equation (5.1) and using the parameters cr_j and pb_t we come to the following objective function for our model:

$$Max \sum_t \left(\sum_i \sum_j \left((p_{t,i,j} - q_i) * cr_{t,i,j} * pb_{t,i} \left(b_{i,t} * Y_{t,i,j} + \sum_{g \in G_i} \sum_h a_{t,g,h} * adr_{t,i,g} * X_{t,i,j,g,h} * rd_{t,i} \right) \right) - \sum_g \sum_h a_{t,g,h} * cpc_{g,h} * Z_{t,g,h} \right) \tag{5.35}$$

Again we can simplify the objective function; we use

$$profitpervisit_{t,i,j} = (p_{t,i,j} - q_i) * cr_{t,i,j} * pb_{i,t} \tag{5.36}$$

and

$$adcosts_{t,g,h} = a_{t,g,h} * cpc_{g,h} \tag{5.37}$$

Then we can write the objective function as

$$Max \sum_t \left(\sum_i \sum_j profitpervisit_{t,i,j} \left(b_{i,t} * Y_{t,i,j} + \sum_{g \in G_i} \sum_h a_{t,g,h} * adr_{t,i,g} * X_{t,i,j,g,h} * rd_{t,i} \right) - \sum_g \sum_h adcosts_{t,g,h} * Z_{t,g,h} \right) \tag{5.38}$$

Again we formulate a capacity constraint:

$$\sum_t \sum_{i \in S_k} \sum_j cr_{t,i,j} * pb_{t,i} * \left(b_{i,t} * Y_{t,i,j} + \sum_{g \in G_i} \sum_h a_{t,g,h} * adr_{t,i,g} * X_{t,i,j,g,h} * rd_{t,i} \right) \le c_k \qquad \forall k \tag{5.39}$$

This capacity constraint is similar to the capacity constraint used in the model currently used, however now we have integrated the PAX booked by visitors attracted via AdWords.

We need two constraints to ensure that in every time period, for every product group and every AdWords group respectively one price or CPC is assigned. These constraints are described as

$$\sum_{j} Y_{t,i,j} \leq 1 \qquad \forall\, t,i \tag{5.40}$$

$$\sum_{h} Z_{t,g,h} \leq 1 \qquad \forall\, t,g \tag{5.41}$$

Besides we need constraints to make sure that if a combination of price and CPC is used, the price and CPC at themselves are used:

$$\sum_{h} X_{t,i,j,g,h} \leq Y_{t,i,j} * rd_{t,i} \qquad \forall\, t,i,j,g \tag{5.42}$$

$$\sum_{j} X_{t,i,j,g,h} \leq Z_{t,g,h} \qquad \forall\, t,i,g,h \tag{5.43}$$

Also in this model all constraints are non-negative:

$$X_{t,i,j,g,h}, Y_{t,i,j}, Z_{t,g,h} \geq 0 \qquad \forall\, t,i,j,g,h \tag{5.44}$$

5.4.3 Discussion

In this section we discuss the application of model proposed in Section 5.3.2 for Sundio. We have implemented the model in AIMMS 3.13; Appendix G shows the text representation of the model in AIMMS. Using CPLEX 12.6, AIMMS is able to solve a problem with 25 time periods, 120 product groups and 11 AdWords groups within 60 seconds. Using only two time periods reduces the time needed to solve the model to approximately 8 seconds. For comparison: the number of AdWords group for the country Turkey is 56, and the number of product groups to optimize is around 70000. We tried to perform a run on a basic desktop pc for a sample dataset with this many AdWords groups and product groups but were not able to solve this problem due to AIMMS was running out of memory.

To estimate the running time for such a large problem we evaluated the running times for solving the models in the simulation study as described in Chapter 7. We calculate a regression based on the number of constraints, variables and nonzeros for all the times that we solved on of the models used. The sample model which we set up with 77700 product groups and 82 AdWords groups has 156094 constraints, 82037665 variables and 404002717 Nonzeros according to the progress windows of AIMMS. When we use the intercept and slopes from the regression based on the simulation study (see Appendix H), we get an estimated running time of 3293 seconds, so almost one hour.

The time needed for solving large countries with many time periods could be a problem when implementing the model; however it is also possible to decrease the number of variables and constraints by solving the model per region instead of per country. Besides the solution for the optimal price and CPC is the only one which can really be used at Sundio since in this model we take the network effects into account. This is the biggest improvement compared to the model discussed in the previous section and makes the model applicable for usage by Sundio.

Another advantage is that the value of the different AdWords group is better valued for AdWords group that attract visitors for multiple products. The value of an advertisement at country level can be much higher than the value of an advertisement at for example product level when the ad ratios for the AdWords group at country level are high; on the other side advertisement groups with a destination URL at a deeper level can be used to manage the number of visits, and therefore the demand, at a lower level.

We also want to notice that with the model discussed it is also possible to set a budget for the expenditures on AdWords. This can be done by using the constraint formulated in (5.42).

$$\sum_{g}\sum_{h} a_{t,g,h} * cpc_{g,h} * Z_{t,g,h} \leq AdWords\ budget_t \qquad \forall\, t \tag{5.45}$$

In the left hand site the total AdWords costs are calculated for every week, these should be smaller or equal to the budget set per week. One can choose to define a budget per week, use the same budget for every week or to set a budget for the remaining booking period. When setting a budget for the remaining booking period, the left hand side should be expanded by taking the sum over all time periods.

5.5 Conclusion

In this chapter we answered the third sub-question formulated in Chapter 2.2: "*How can the effects of the decisions in pricing and advertising be combined in a model?*" We started with introducing a continuous model for a simplified case where there is one product group, one AdWords group, no capacity restriction and one price and CPC to set for the remaining period. We extended the model first by taking into account the capacity restriction. Then we added the possibility to assign different prices and CPCs per time period. Finally we added the possibility to optimize the price of multiple product groups and the CPC of multiple AdWords groups.

The final model is formulated as

$$Max \sum_{t}\left(\sum_{i}\sum_{j} profitpervisit_{t,i,j}\left(b_{i,t} * Y_{t,i,j} + \sum_{g\in G_i}\sum_{h} a_{t,g,h} * adr_{t,i,g} * X_{t,i,j,g,h} * rd_{t,i}\right) - \sum_{g}\sum_{h} adcosts_{t,g,h} * Z_{t,g,h}\right)$$

$Subject\ to$

$$\sum_{t}\sum_{i\in S_k}\sum_{j} cr_{t,i,j} * pb_{t,i} * \left(b_{i,t} * Y_{t,i,j} + \sum_{g\in G_i}\sum_{h} a_{t,g,h} * adr_{t,i,g} * X_{t,i,j,g,h} * rd_{t,i} \right) \leq c_k \qquad \forall k$$

$$\sum_{j} Y_{t,i,j} \leq 1 \qquad \forall\, t,i$$

$$\sum_{h} Z_{t,g,h} \leq 1 \qquad \forall\, t,g$$

$$\sum_{h} X_{t,i,j,g,h} \leq Y_{t,i,j} * rd_{t,i} \qquad \forall\, t,i,j,g$$

$$\sum_{j} X_{t,i,j,g,h} \leq Z_{t,g,h} \qquad \forall\, t,i,g,h$$

$$X_{t,i,j,g,h}, Y_{t,i,j}, Z_{t,g,h} \geq 0 \qquad \forall\, t,i,j,g,h$$

The main benefit of this model compared to the current model used is that it combines the effect of changing the price and changing the CPC in one model. Besides the optimization of the CPCs set in AdWords improves by taking into account the capacity of a product. By using multiple time periods the model also takes expected price changes into account.

In the next chapter we discuss how a forecast can be made for the different parameters which are added to the model: baseline visits, visits from Google AdWords, Ad ratio, conversion rate and PAX per booking. In Chapter 7 we use a simulated test case to see how the model performs compared to the model currently used.

6 Forecasting

In the previous chapter we proposed a new model to use at Sundio. For this new model different parameters have to be forecasted. In this chapter we answer the fourth research question: "*How can we use the data available regarding pricing and advertising decisions to forecast the effects of pricing and advertising on attraction and conversion?*" To answer this question we use the different forecasting methods discussed in our literature review in Chapter 4. Per parameter which has to be forecasted we compare the different forecasting methods based on the measures of forecast accuracy and then determine which forecasting method should be used for the parameter.

For these analyses we use data in the time period from the fourth of October 2010 till the 29th of September 2013. This period starts with the first day of week 40 in 2010, and ends at the last day of week 39 in 2013. In this analysis we choose to aggregate the data over the weeks because otherwise the data is too specific, this would make it more difficult to make a correct forecast. For the analysis we consider the data from week 40 till week 39 as the data of one booking period, where the period from week 40, 2010 till week 39, 2011 is referred to as the booking period of the year 2011. We use the data of the booking periods 2011 and 2012 to make a forecast for booking period 2013. We compare the forecast for 2013 with the actual data of 2013 for calculating the forecasting accuracy. Based on the forecasting accuracies of the different forecasting methods we give an advice about which method to use for each parameter.

In our analysis we focus on the data related to Sunweb Zomer, the largest brand of Sundio offering holidays with departure dates mostly in the period from March till August. We gather the data used from the data warehouse of Sundio and directly from Google Analytics/AdWords. We do not take cancelled bookings into account. In this analysis we only consider the advertisement campaigns for non-branded keywords. This means that we focus on keywords not containing the word Sunweb. The reason for not taking into account the branded keywords is that these customers are already focused on Sunweb and would probably also have visit the site without the advertisement being visible next to the search results. Besides branded advertisement cannot be compared to non-branded advertisement since the costs are much lower; the reason for this is that Google does not allow competitors to advertise on keywords containing a brand.

In Section 6.1-6.5 we respectively discuss forecasting the baseline visits, the number of impressions in AdWords, the click through rate in AdWords, the ad ratio, the conversion rate, and the average number of PAX per booking. We conclude this chapter in Section 6.6 with a summary of the results.

6.1 Baseline visits

We start with making a forecast for the number of baseline visits for a product group. We define the number of baseline visits as the number of visits at the accommodation level since this is the level where a visitor sees the prices of a product and can start making a booking. Note that for the optimization model there are many product groups, however for forecasting the number of visits we do not have to make a distinction on the duration and the departure date of a product group. The reason is that the number of baseline visits for products groups for which only the duration and the departure date are different, have the same webpage and therefore the same number of baseline visits.

For forecasting the number of baseline visits we consider the different accommodation groups related to Turkey. With an accommodation group we refer to a group of accommodations that are in the same product group. We use this country for our analysis since this is the most booked destination country of Sunweb Zomer. This country has 22 different accommodation groups when not taking into account the departure date and the duration. We analysed three accommodation groups with respectively 1, 11 & 25 accommodations to get an insight in which forecasting method performs the best for product groups of different sizes. An overview of the number of visitors is given in Figure 12.

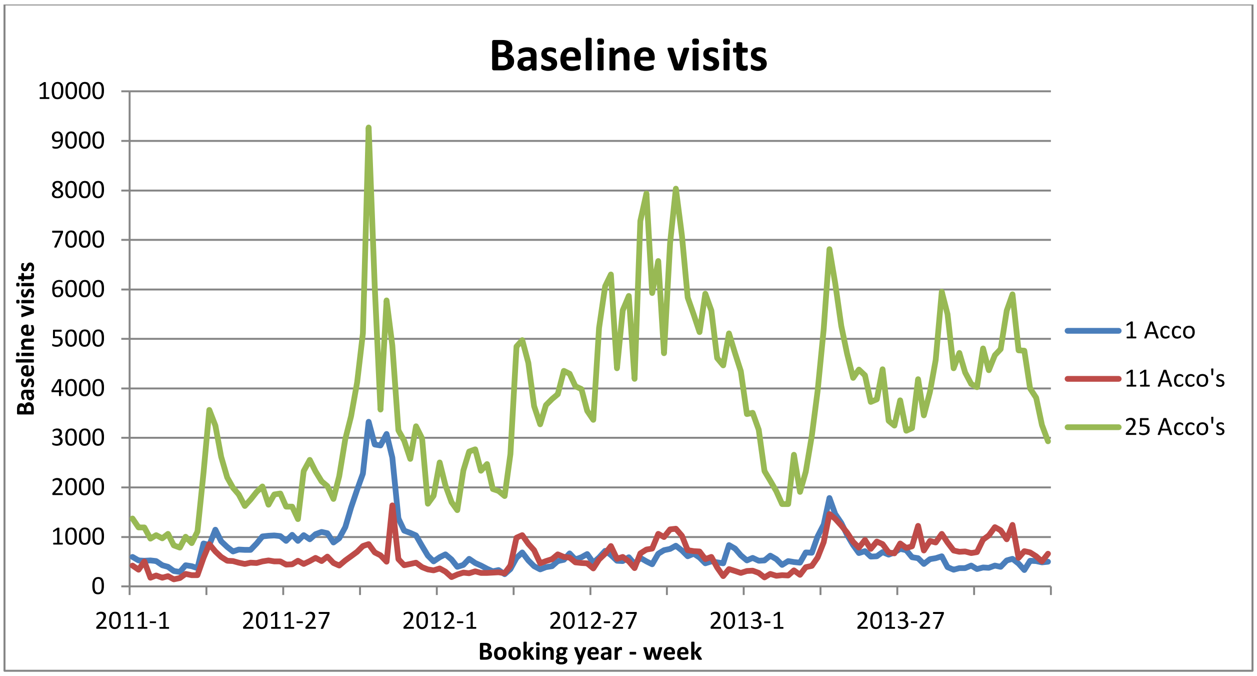

Figure 12: Overview number of baseline visits for different product groups

For the number of baseline visits used in the API model we need a forecast per single week. We propose the following forecast methods:

1. Last week: $\hat{y}_t = y_{t-1}$
2. Same week last year: $\hat{y}_t = y_{t-52}$
3. Average same week last two years: $\hat{y}_t = \frac{y_{t-52}+y_{t-104}}{2}$
4. Calculate ratio between last five weeks this year and previous year and multiply same week last year with this ratio: $\hat{y}_t = \frac{\sum_{i=1}^{5} y_{t-i}}{\sum_{i=1}^{5} y_{t-i-52}} * y_{t-52}$
5. Additive Holt Winters' method as described in Section 4.2.1 with a damped trend and a length for seasonality of 52 week.
6. Multiplicative Holt Winters' method as described in Section 4.2.1 with a damped trend and a length for seasonality of 52 week.

For the different methods, except for using last week's value, we also consider smoothing the forecasted value of 3, 5 or 7 weeks. For the first method smoothing makes no sense because every forecasted value is equal to the value of last week. For the first three methods the forecast for a specific week does not change during the booking year. For method 4, 5 and 6 the forecasts change during the booking period; for the Holt Winters' methods new parameters are calculated every week. The smoothing parameters for the Holt Winters' method are calculated every week with the new information included. The parameters are chosen such that the MAPE value over the period of the last 52 weeks is minimized. For the Holt Winters' method we use seasons of one week; using a short period the peaks around the end of the calendar year and in the summer period would not be visible. On the other hand using seasons of one month would result in flatting the effects of the peak periods out, which is a problem since we want to optimize on a weekly basis. A disadvantage of using one of the Holt Winters' methods is that they have a large computation time since the parameters used have to be optimized for every product group. One could choose to use the same smoothing parameters for every product group; however this probably results in worse forecasting results since different product groups have different characteristics.

Table 10 (Appendix I) shows the MAPE values for using the different methods proposed for forecasting the number of baseline visits for the different product groups considered. These MAPE-values are based on $\sum_{n=1}^{52} n$=1378 forecasted values; for every week in booking year 2013 a forecast is made for the remaining weeks. Using this method, the MAPE value of the last week in the booking period has a higher weight; however this is justified by taking into account that the forecast of this week is used more times. For example the forecast for the second week in the booking period is only used when running the model at the start of the first and second week; on the other side the forecast value for the last week is used in every optimization in the booking period.

Using the MAPE values as a measure of accuracy for forecasting we see that the simple forecasting methods outperform the more complex Holt Winters' method in most cases. A reason for this could be that the peaks in the number of baseline visits differ per booking year and thereby disrupt the forecast of the Holt Winters' method. The other methods are not based on the complete history and therefore are less disrupted by the irregular peaks. Based on the MAPE values we conclude that forecasting method 2, using the number of baseline visits in the same week last year, smoothed over 7 weeks, has the best overall performance.

In Section 5.4.3 we also mentioned that it is possible to use two time periods for the Integrated Pricing and Advertising model. Using two time periods means that we use one small time period, for example for the next week, and one large time period for the remaining period. Therefore we also want to know which forecasting method is good in predicting only the next week. Table 11 (Appendix I) shows the MAPE-values for the different methods for forecasting only one week in advance.

Based on the MAPE values as shown in Table 11 (Appendix I) we conclude that using the number of baseline visits in the last weeks is the best method for forecasting the number of visits for the next week. Comparing Table 10 and Table 11 we also see that the forecasting accuracy of the Holt-Winters' methods are much better when only forecasting the next week. This suggests that these methods are not suitable for forecasting many time periods in advance.

We also need to know which forecasting method to use to forecast the number of visitors in the remaining time periods. Note that since there are departures in most weeks, the number of remaining time periods differs per product group. For example when we consider the start of week 4, a product group with a departure in week 8 needs a forecast for week 4 and a forecast for the remaining weeks (week 5, 6, 7 and 8) which differs from a product group with a departure date in week 12. Therefore for comparing the different forecasting methods for forecasting the number of baseline visits in the remaining period, we made a forecast for week in booking year 2013 and for every possible number of weeks in the remaining booking period. Table 12 (Appendix I) shows the forecasting accuracy (MAPE) for forecasting the number of baseline visits in the remaining period.

Based on Table 12 we advise to use forecasting method 2, using the number of baseline visits in the same week last year, for forecasting the number of baseline visits in the remaining booking period. This forecasting method is the most accurate in most cases, only for the accommodation group with 25 accommodations it performs a little worse than forecasting method 3; however method 2 seems to be more robust.

One could try to improve the forecast by taking more variables into account, for example how many and which regions of the Netherlands do have holidays or based on the weather. Based on the limited time available for this research we consider this to be outside the scope of this research and want to suggest this for further research.

6.2 Visits from Google AdWords

The next parameter that we want to forecast is the number of visits attracted via an ad group in Google AdWords. To forecast this value we first make a forecast of the number of impressions and a forecast of the click through rate. By multiplying these forecasts, the number of clicks on advertisements from the ad groups can be calculated. The number of impressions for an ad group in Google AdWords is defined as the number of times that an advertisement from a specific ad group is shown. An ad group is defined as a group of keywords with advertisements with the same destination URL. An exception is made for ad groups at the accommodation level. Here we aggregate all the advertisements for products within the same product group as one ad group because these visitors are considered to visit the same product group.

For the analysis of which forecasting method to use for forecasting the impressions we use one AdWords group at each web site level. This means that we consider an ad group at country level (Turkey), region level (Turkish Riviera), city level (Alanya) and product group level (we use the same product group as used in the previous section with 25 accommodations).

Since Google only shows a limited number of advertisements next to the search results, the number of impressions depends on the position assigned by Google. For example, when Google only shows three advertisements and the fourth position is assigned to an advertisement, Google does not show this advertisement and therefore there is no impression for this advertisement. From Google AdWords we retrieved the percentages indicating the number of impressions received divided by the estimated number of impressions the advertisement was eligible to receive; we call this the impression share. An advertisement is eligible to be shown next to the search result when a user uses related key words, however due to the limited number of advertisement positions being available an advertisement with a relatively low bid is not shown every time. The data related to the impression shares is available for the period starting at week 5 in booking year 2013; this means we have 48 weeks of data available.

We expect that a higher CPC results in a higher impression share because Google AdWords is more willing to show the advertisement. Figure 13 shows the scatterplots of the impression share versus the CPC for the different AdWords groups considered.

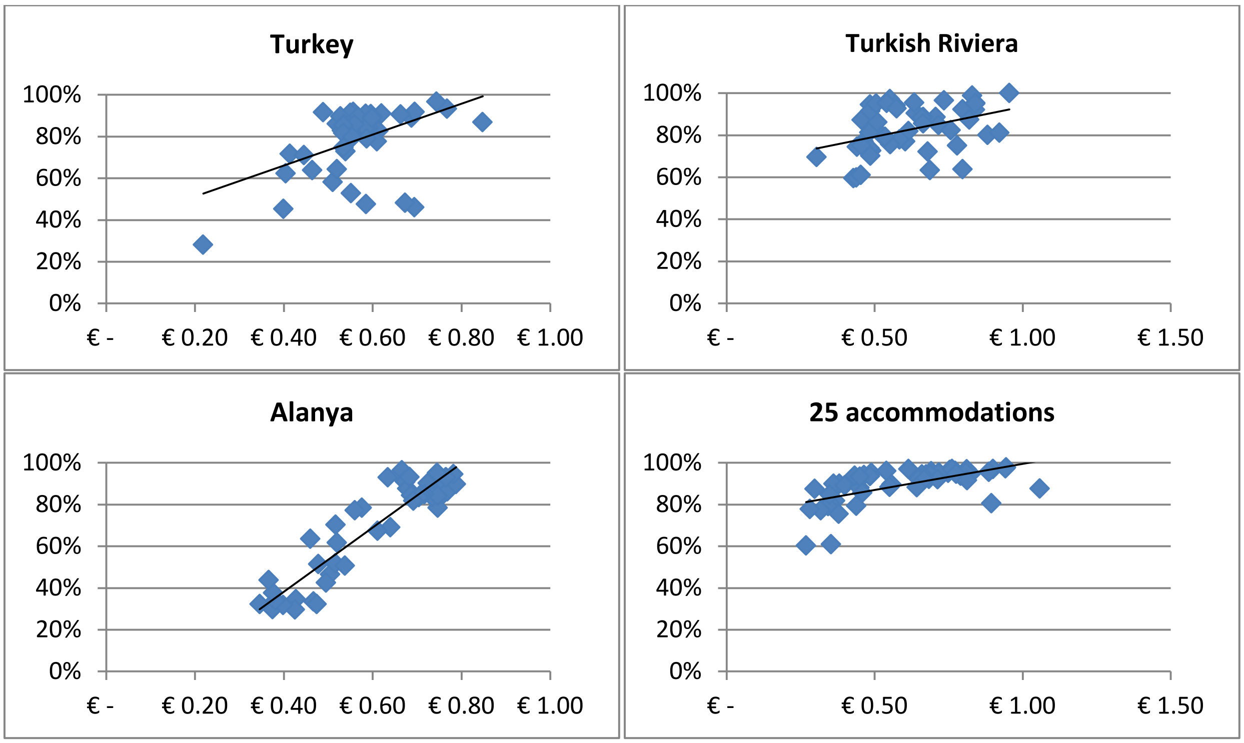

Figure 13: Scatterplots of the impression share versus the Cost Per Click for different AdWords groups

As expected we see a positive relation between the CPC and the impression share. The related ANOVA tables are shown in Appendix J. The ANOVA tables show a high significance for the different regression analysis ($\alpha < .05$ in all cases); the adjusted R square value for the different regressions ranges from .14 till .85. The adjusted R square value is an indication for how much of the variance in the output variable is explained by the input variable. For different AdWords groups the forecasting accuracy is different because of the number of keywords into the AdWords group and the diversity of the different keywords. For example the R square value for Turkey is low because the variety between the key words is very high (e.g. “Turkije” and “Vakantie turkije met kinderkorting”); the data for the Turkish Riviera is based on a total of approximately 50k impressions while for the city Alanya over 1 million impressions where tracked.

Now that we know that there is a significant relation between the impression share and the CPC we have to determine how to forecast the impression share in the future based on the CPC set. Since the position in Google AdWords, and therefore also the impression share, also depends on the bids of competitors, we want to know how many weeks of data we should use to determine the parameters of the regression used for forecasting the impression share.

For different numbers of weeks taken into account we determine the accuracy for forecasting next week's impression share. Besides we used two different methods for using the regression. The first method uses the intercept and beta as calculated with the regression. This method can be formulized as

$$y_{t+1} = \alpha + \beta * CPC_{t+1} \tag{6.1}$$

The other method uses the effect calculated with the regression and applies it to the change in CPC compared to last week and the impression share of last week. This method can be written as

$$y_{t+1} = y_t + \beta * (CPC_{t+1} - CPC_t) \tag{6.2}$$

First we calculate the MAPE value for using both methods and taking different numbers of weeks into account for the different AdWords groups. Then we calculate a weighted average of the MAPE based on the total costs for the different AdWords groups used; the MAPE values are weighted based on the total costs of an AdWords groups. Since we only have 47 weeks of impression shares available the longest time window used for calculating the regression that we considered is 27 weeks; therefore the MAPE values are based on forecasting the latest 20 weeks from the data available. The MAPE values for using different weeks of data and two different methods of applying the regression found is shown in Figure 14.

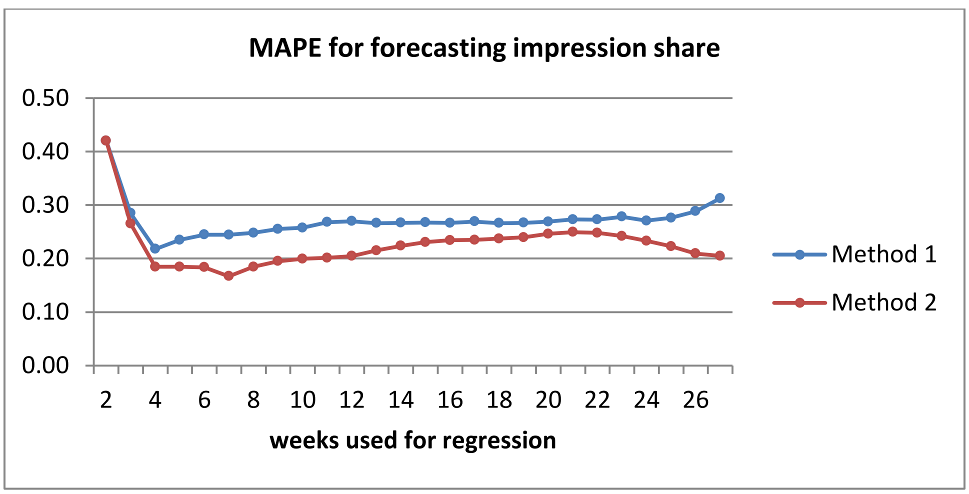

Figure 14: Weighted MAPE values for forecasting the impression share using two different methods of applying the regression and different numbers of weeks taken into account

In Figure 14 we see that using method 2, applying the regression on last week's value instead of using the intercept, is the best method for making a forecast of the impression share of next week. The best number of weeks to use for the regression is seven weeks; this results in an MAPE value of 0.17.

Based on the regression analysis performed for the impression share versus the CPC we conclude that the number of impressions received depends on the CPC used. Therefore we want to normalize the number of impressions for the different AdWords groups based on the relation found between the impression share and the CPC for the time period for which the impression share is not known. After normalizing the number of impressions to get the number of eligible impressions we analyse which forecast method to use to forecast the number of eligible impressions. When the number of eligible impressions is forecasted we use the alpha and beta found from the regression analysis in combination with the CPC set to forecast the number of impressions.

An overview of the number of visitors is given in Figure 15. We see that the number of eligible impressions differs a lot for the different AdWords groups. However we see some similarities in the patterns. We expect that seasonality influences the number of eligible impressions since in some periods people are more willing to perform a holiday related search request via Google than in other periods.

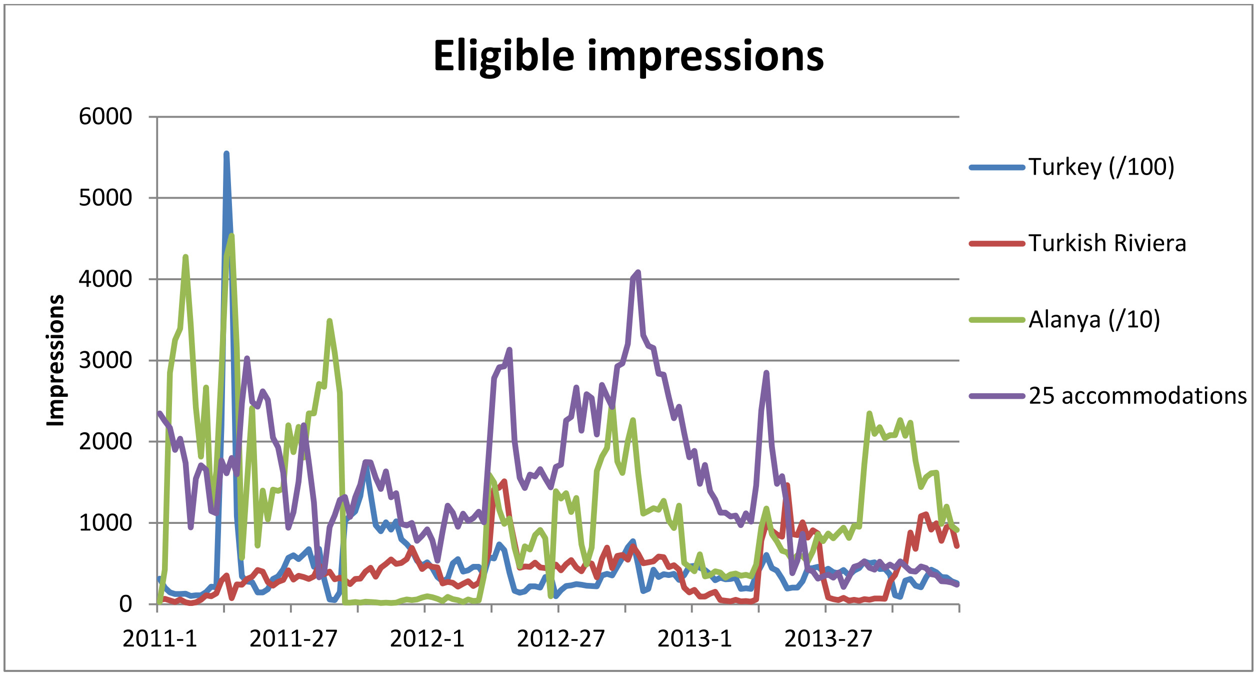

Figure 15: Overview number of eligible impressions for different AdWords groups

In our analysis for forecasting the eligible number of impressions we use the same forecasting methods as used in the previous section for forecasting the baseline visits. Table 13 shows the MAPE values for using the different methods proposed for forecasting the number of eligible impressions in Google AdWords.

Analogous to the forecasting accuracies for forecasting the number of baseline visits we see that the naïve forecasting methods outperform the Holt Winters' methods in most cases. Only for forecasting the number of eligible visits for the Google AdWords with a destination page for the accommodations the multiplicative Holt Winters' methods has the lowest MAPE value, however this value is comparable to the MAPE value for using last week's value as a forecasting method. Based on the MAPE values shown in Table 13 (Appendix I) we conclude that using last week's value as a forecasting for the remaining weeks is the best method to forecast the eligible number of impressions.

As mentioned in the previous section we also want to know how the forecasting models perform when using two time periods. Table 14 (Appendix I) shows how the different forecasting methods perform when forecasting the number of eligible impressions in Google AdWords for the different AdWords Groups considered.

Also for forecasting the number of eligible impressions in the next week, the naïve forecasting methods have the best accuracy. We see that using the number of eligible impressions in the last week as a forecast for the next week has the best performance in all cases. The same forecasting method is used in the current method for forecasting the number of impressions. Based on the MAPE values shown in Table 14 we advise to also use this method for forecasting the number of eligible impressions in the next week.

As for the baseline visits, we also compare how the different forecasting methods perform when making a forecast for the total number of eligible impressions in the remaining weeks. Table 15 (Appendix I) shows the MAPE values for the different methods when making this forecast. Again we see that the naïve forecasting methods outperform the Holt Winters' methods. Based on the MAPE values shown in Table 15 we advise to use the number of eligble impressions in the previous week as a forecast for the remaining weeks since this method seems to be the most robust.

Now that we know how to forecast the number of impressions for an AdWords group we need to determine how the click through rate for an AdWords group can be forecasted in order to be able to calculate the expected number of visits from AdWords. The click through rate is defined as the number of people that clicked on an advertisement of the AdWords group divided by the number of impressions that the AdWords group had. Since the CTR depends on the CPC chosen for an AdWords group we use a linear regression to estimate the effect of changing the CPC on the CTR. Figure 16 shows the scatterplots of the impression share versus the CPC for the different AdWords groups considered.

As expected we see a positive relation between the CPC and the CTR. The related ANOVA tables are shown in Appendix K. The ANOVA tables show a high significance for the different regression analysis ($\alpha < .05$ in all cases); however the adjusted R square value for the different regressions ranges from .41 till .59.

Since the CPC of competitors, and therefore the CTR, can change over time we also analyse how many weeks of data should be taken into account for calculating the regressions for the relation between the CTR and the CPC. We analyse this in the same way as for the impression share and again use the two methods proposed for using the regression. Since we have more data available for the CTR we now consider looking back 2 till 52 weeks. We use the latest 52 weeks of the data available to calculate the MAPE. The MAPE values for using different weeks of data and two different methods of applying the regression found is shown in Figure 17.

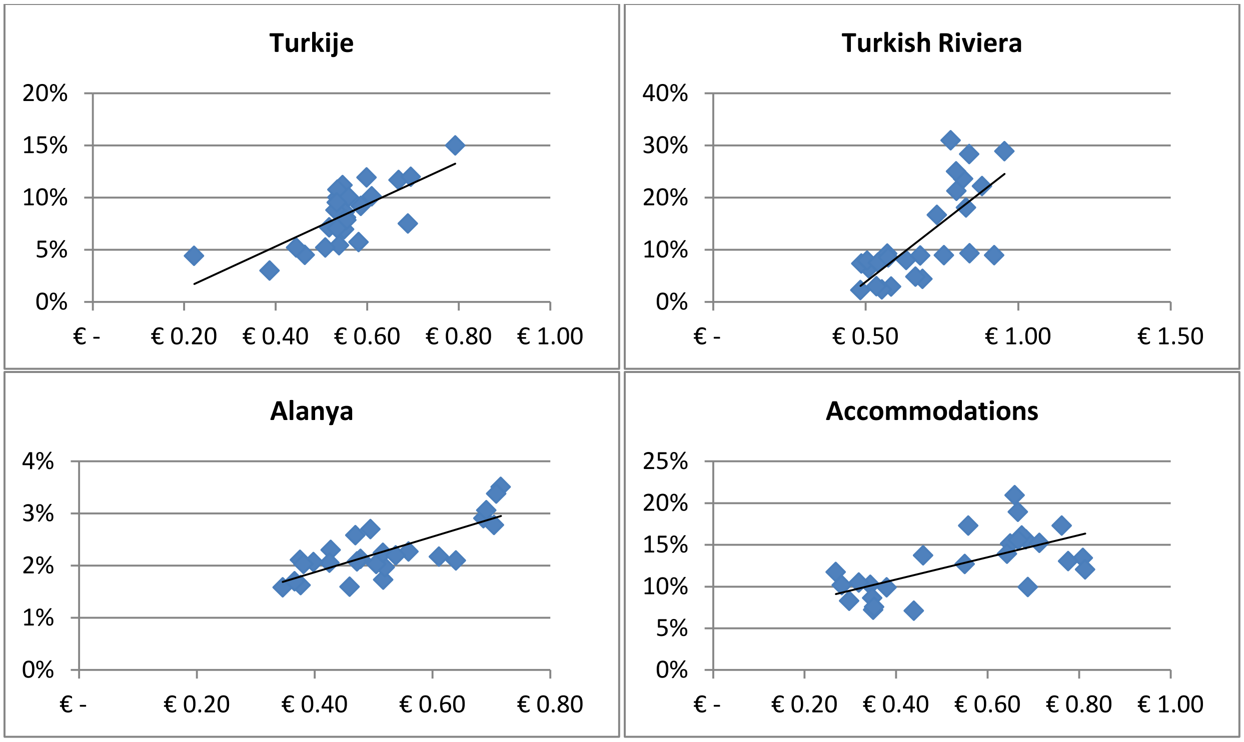

Figure 16: Scatterplots of the Click Through Rate versus the Cost Per Click for different AdWords groups

Again we see that using method 2, not the intercept of the regression but the value of last week, is better than method 1. Using the last five weeks for the regression results in the lowest MAPE value (0.17). Therefore we advise to forecast the effect of the CPC based on the regression for the last five weeks and apply it on the CTR of last week.

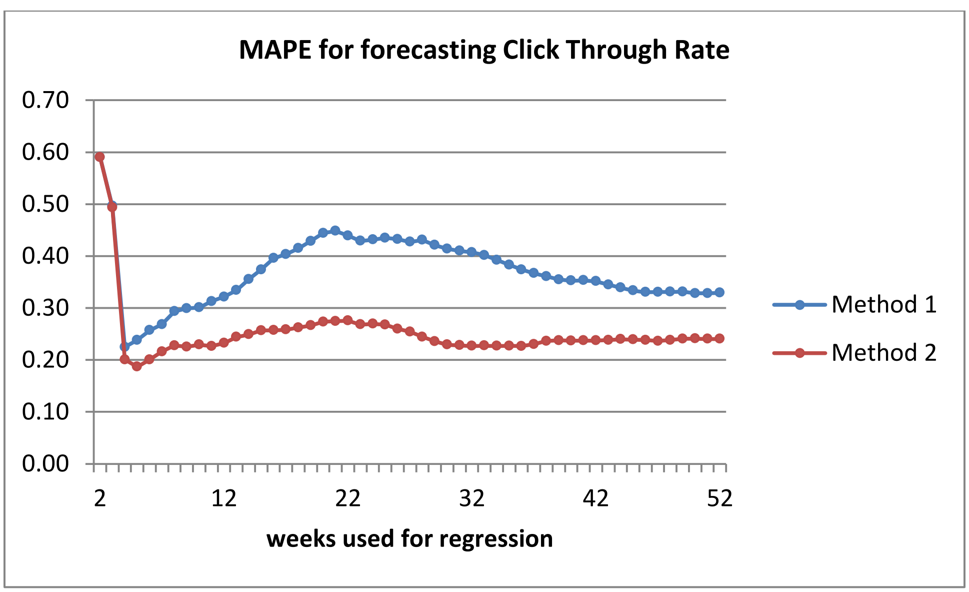

Figure 17: MAPE values for forecasting the Click Through Rate using two different methods of applying the regression and different numbers of weeks taken into account

6.3 Ad ratio

In the previous section we explained how the number of visitors attracted by each ad group can be calculated. In this section we discuss how the ad ratio between a product group and an AdWords group can be calculated. The ad ratio indicates what ratio of people who click on an advertisement those also visit the webpage of an accommodation where they can buy a product. For example when 100 people enter the website via an advertisement with the destination URL of a country (e.g. zomer.sunweb.nl/turkije/) and ten of them visit the accommodation page of a product group, the ad ratio between that destination URL and the product group is 0.10. Since these data are not directly available we use the ratio of baseline visits at the destination URL of the advertisement that visits a webpage at the accommodation level of the product group as an estimate for the ad ratio.

For our analyses we use the baseline visits at country level (Turkey), region level (Turkish Riviera), and city level (Alanya) and compare them with the number of visits at an accommodation level (we use the same accommodation group as used in the previous section). Off course, the ad ratio for the AdWords with a destination URL at the accommodation level is always 1 since all the visitors for these advertisements visit the page of the related accommodation group. Figure 18 provides an overview of the visit ratios related to these numbers of visitors. We clearly see that the Ad ratio for deeper website levels is higher. This is also what we expected since the visitor is closer to the pages related to the apartments in his search for a holiday when he is at a deeper level at the website.

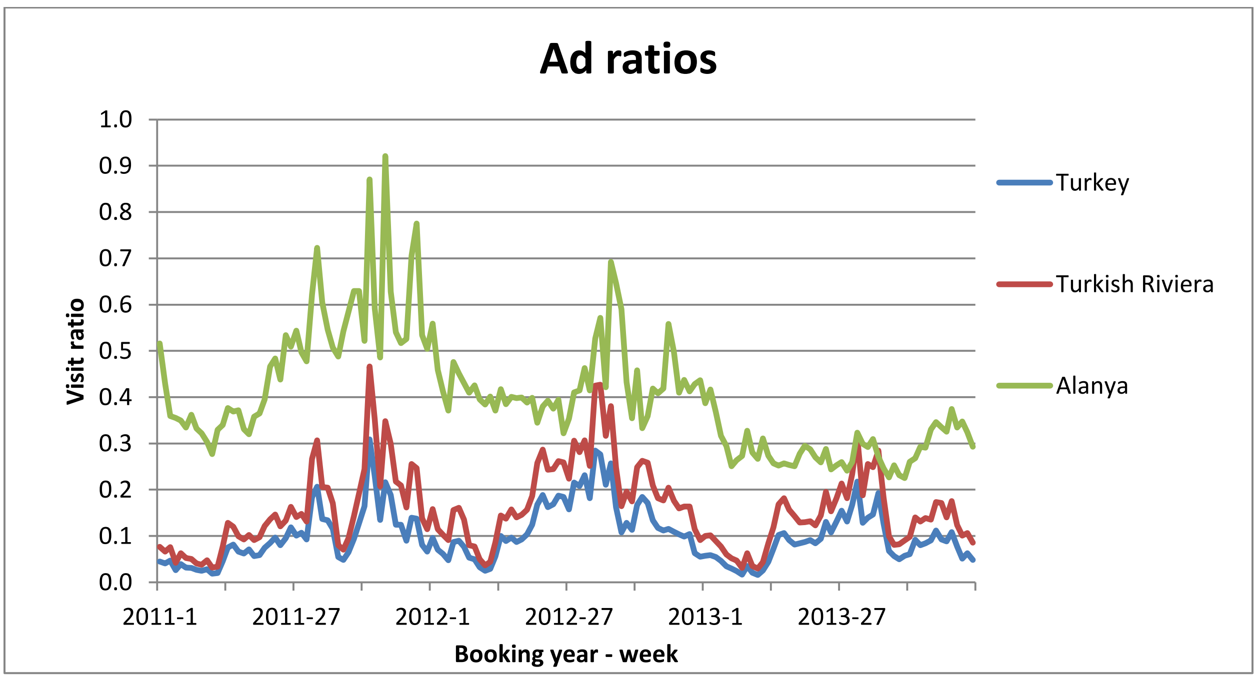

Figure 18: Overview of the ad ratios of visitors of different destination URLs to the accommodation pages

In the same way as in the previous two sections we compare the accuracy of the different forecasting methods mentioned in Section 6.1 to see which method should be used to forecast the ad ratios. Table 16 (Appendix I) shows how the different forecasting methods perform when forecasting the ratio between the numbers of visitors at the destination URL that visits the accommodation page.

In our opinion method 4, calculating the ratio based on the ratio between this year and previous year in the last five weeks, is the best forecasting method for forecasting the ratio between the numbers of visitors at the destination URL that visits the accommodation page. Again we also investigate what forecasting method is the best method to forecast the ratio for only the upcoming week. Table 17 (Appendix I) shows the MAPE values for the different methods when forecasting only one week. We see that in this case using the ratio of last weeks has the best accuracy.

In our opinion method 4, calculating the ratio based on the ratio between this year and previous year in the last five weeks, is the best forecasting method for forecasting the ratio between the numbers of visitors at the destination URL that visits the accommodation page. Again we also investigate what forecasting method is the best method to forecast the ratio for only the upcoming week. Table 17 (Appendix I) shows the MAPE values for the different methods when forecasting only one week. We see that in this case using the ratio of last weeks has the best accuracy.

For the remaining period in case of using two time periods we test which average of the forecasted ratios for the remaining booking period is closest to the real average of the ratios. Table 18 (Appendix I) displays the MAPE values for forecasting the ratio. Again we see that the best method for forecasting the ratio between the numbers of visitors at the destination URL that visits the accommodation page in the remaining booking period is method 4. Therefore we advise to use this method when forecasting the average ratio for the remaining booking period.

Again we see that the best method for forecasting the ratio between the numbers of visitors at the destination URL that visits the accommodation page in the remaining booking period is method 4. Therefore we advise to use this method when forecasting the average ratio for the remaining booking period.

Since there is no data available about which pages are visited at the website by visitors attracted via Google AdWords we do not really know whether the calculate number are a good representation of this data. Although, in this section we analysed which methods should be used when forecasting an estimation of the ad ratio.

6.4 Conversion rate

The next parameter for which we analyse which forecast method to use is the conversion rate. The conversion rate is defined as the number of bookings made divided by the number of visits. For our model we calculate the conversion rate per product group. For our analysis we consider again three product groups, we use the same product groups as used in Section 6.1.

At a product group level the number of visitors and the number of bookings is low. The three product groups considered have on average respectively 748, 3596 and 596 visits per week. This means that when the average conversion rate is for example around 0.001, there probably are some weeks in which there are no bookings. This makes it difficult to estimate the “real” probability that a visitor converts into a customer making a booking. Therefore the forecast should be made at a level at which a sufficient number of bookings is made per week. In this section we use the country level as the level at which we make a forecast for the conversion rate. This means that for example a forecast for the conversion rate of a product group with destination region Turkish Riviera is based on the conversion rate for Turkey.

First we want to know how the conversion rate develops over time. We use the expected development of the conversion rate in the same way as the booking curve is used currently. The expected development of the conversion rate progress indicates how many bookings per visitor are expected over time and the booking curve indicates how many PAX are expected.

In this section we explain how the conversion rate progress for the different accommodation groups can be constructed similar to how the booking curve is constructed. We use the construction of the conversion rate progress for departure week 33 (week 46 in booking year) as an example.

First the conversion rate progress at the country level should be forecasted. This can be done by taking the average of the conversion rate in the previous two years for the same weeks. Figure 19 shows the conversion rate progress of booking year 2011, 2012 and 2013 for packets to Turkey and the forecast for 2013 based on 2011 and 2012. The forecasted conversion rate looks like a good estimation of the real conversion rate. The MAPE value of this forecast is 0.22; using just last year's conversion rate results in an MAPE value of 0.33.

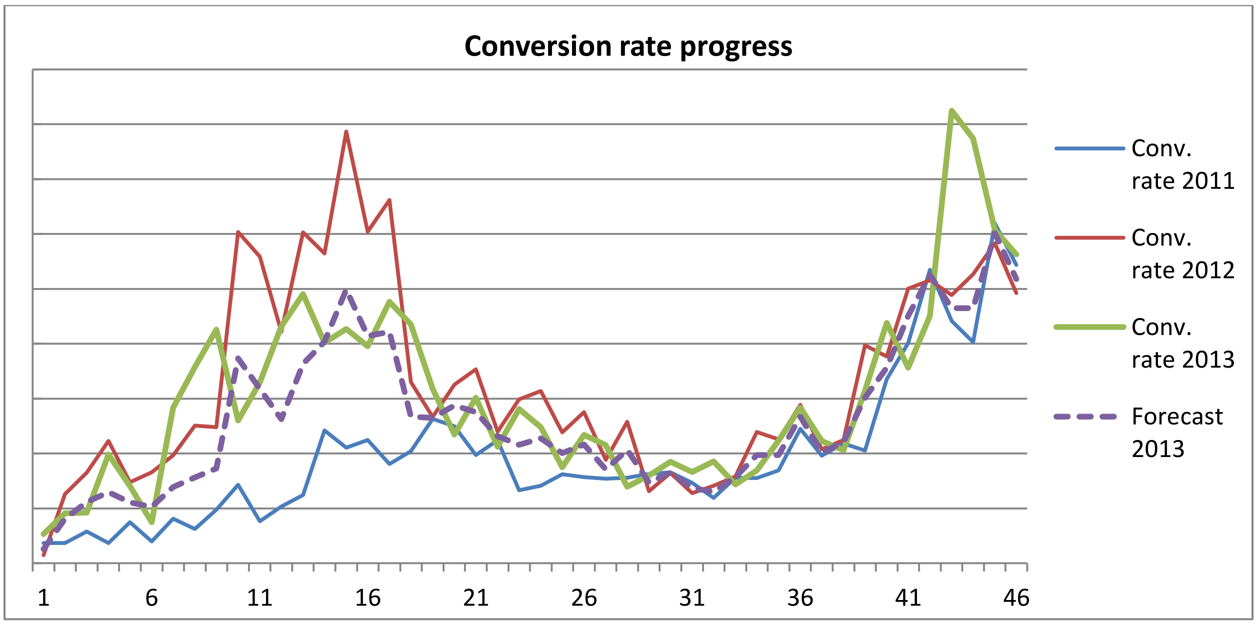

Figure 19: Conversion rate progress for booking year 2011-2013 and the forecast for 2013 for destination country Turkey and departure week 33 (y-as labels removed because of confidentiality)

The forecast of the conversion rate progress for Turkey can be used to make a forecast of the conversion rate for a product group. In the current situation the booking curve is in percentages and uses the current number of PAX to make a forecast. For the conversion rate we propose to use the conversion rate over the preceding weeks to level the conversion rate progress. For example when the conversion rate for Turkey in the preceding weeks was .002, and for the product group it was 0.004, we level the conversion rate for the product group by multiplying it with .004/.002=2. This is based on the assumption that the ratio between the conversion rate of a product group and the conversion rate of a country will have the same proportion in the remaining weeks as in the preceding weeks. In the left graph of Figure 20 we see this ratio for the remaining and preceding weeks; the right graph in this figure shows how the conversion rate of Turkey levelled based on the cumulative ratio of the preceding weeks matches the conversion rate of the product group.

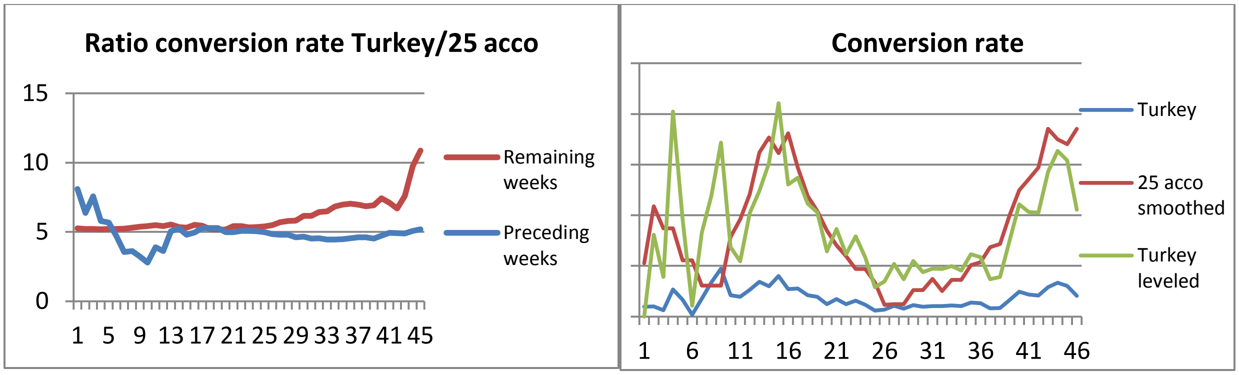

Figure 20: Ratio between conversion rate Turkey and the Accommodation group (left) and the conversion rate per week (right) for destination country Turkey and departure week 33 and accommodation group with 25 accommodations (y-as labels removed because of confidentiality)

In the left graph of Figure 20 we see that the conversion rate of the accommodation group is much higher than the conversion rate for whole Turkey. In the right graph we levelled the conversion rate progress of Turkey based on the conversion rate in the preceding weeks. We see that the levelled conversion rate progress, our forecast for the accommodation group, matches the conversion rate of the accommodation group quite well.

The forecasted conversion rate for the accommodation group is used as the baseline conversion rate. This conversion rate applies when using the reference price. For the reference price we use the same calculations as in the current situation. When the price differs from the reference price we use the price elasticity to calculate the conversion rate corresponding to this price.

In the current situation the price elasticity is defined as the percentage of change in demand divided by the percentage of change in price. For example if the price of a product changes from € 400 to € 440 (+10%) and demand changes from 80 PAX to 60 PAX (-25%), the price elasticity is -25/10=-2.5. In the new situation we assume that the price influences the demand in terms of conversion rate instead of number of PAX booked. Therefore in the new situation we define the price elasticity as the percentage of change in conversion rate divided by the percentage of change in price.

For the new situation we propose to use to use the same way of calculating the price elasticity as in the current situation (as explained in Appendix A). However, in the calculations we change the demand in PAX by the conversion rate. This means that when for example the price elasticity is -2.5 and the price is increased with 10%, the conversion rate would decrease with 25%. We calculated the price elasticity over time for Turkey and saw that the price elasticity ranges from -1 till -3.1. Using the same data and the current way of calculating the price elasticity we see that the price elasticity ranges from -1 till -3.9. Since we calculate the price elasticity in the same way as in the current situation we also choose to calculate the reference price in the same way.

6.5 PAX per booking

The last parameter for which we determine the best forecast method is the average number of PAX per booking. Since the number of bookings for an accommodation group is limited, we want to forecast the number of PAX per booking at the region level. For our analysis we consider the average PAX per booking in Turkey.

First we evaluate whether we should look at the average PAX per booking in a booking week or at the average PAX per booking for a number of weeks before departure. To evaluate this we order the bookings available in the booking data by the number of weeks before departure and by the week number of the booking week. We calculate the variance of the number of PAX per booking within these weeks and calculate for both methods the average variance weighted by the number of bookings made in a week. When using the average number of PAX per booking in a booking week the weighted average variance within a week is 2.15. Making the same calculation for the numbers of weeks before departure we calculate a weighted average variance within a week of 2.09. Therefore we choose to determine the average PAX per booking for each number of weeks before departure.

The average PAX per booking depends not only on the number of weeks in advance, but also on the duration of the holiday. Figure 21 shows the average number of PAX per booking over the weeks before departure for different travel lengths. We see that for holidays with a length of eight or less days the average number of PAX per booking is higher than for holidays with a length of more than 15 days. Besides we see that for holidays with a length of eight or less days the average number of PAX per booking is declining over time, while for holidays with a travel length of more than 15 days the average number of PAX is more constant over time.

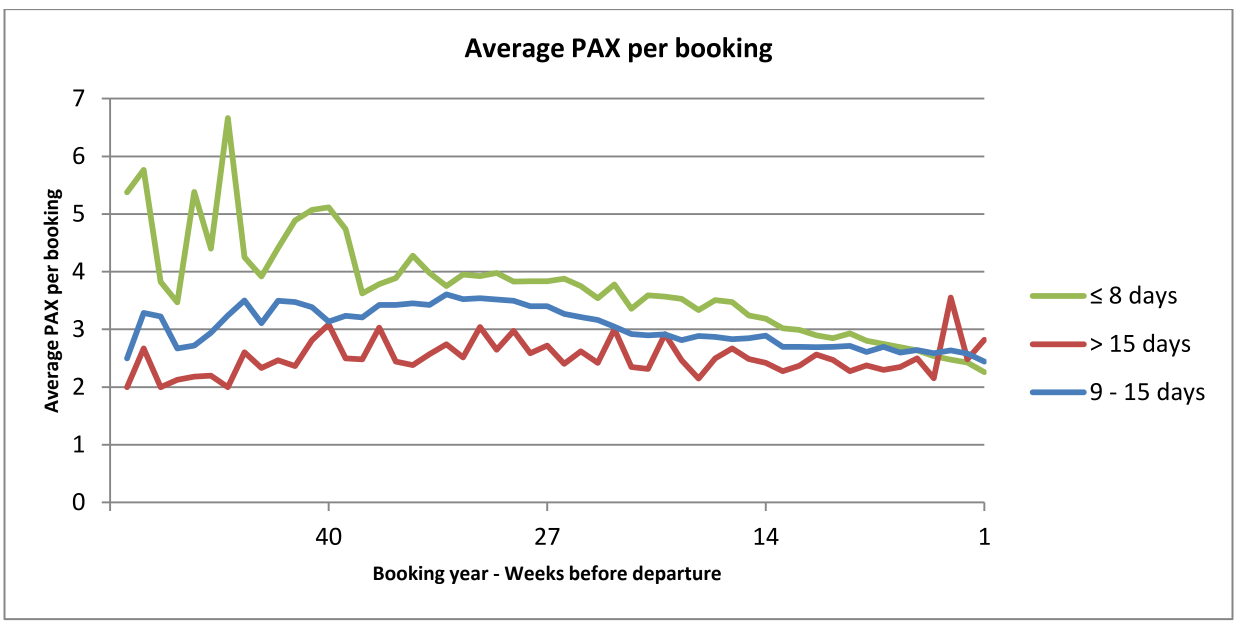

Figure 21: Average number of PAX per booking for different travel lengths

We also see that more than 40 weeks before departure the average number of PAX per booking is fluctuating more than closer to departure; the reason for this is that more than 40 weeks before departure the number of bookings is very small. This also supports that a high number of bookings is needed in order to make a good estimate of the average number of PAX per booking.

For analysing how to forecast the average number of PAX per booking we consider the same forecasting methods as in the previous sections. Table 19 (Appendix I) shows the MAPE values for the different methods when forecasting the average PAX per booking. Based on the MAPE values shown in Table 19 we conclude that method 2, using the same average PAX per booking as in the same week last year, results in the best forecast.

Again, we also analyse how the methods perform when forecasting the next week. The related MAPE values for forecasting only the next week are shown in

Table 20 (Appendix I).

For forecasting the average number of PAX per booking in the next week and in the remaining period, the forecasting method of using the number of PAX per booking last year has the highest accuracy. Based on the MAPE values we conclude that this method should be used when forecasting the average number of PAX per booking.

6.6 Conclusions

In this chapter we discussed how the different parameters that are used in the Integrated Pricing and Advertising model can be forecasted. When comparing the accuracy of the different forecasting methods used, *we noticed that in all the naïve methods result in a higher forecasting accuracy than the additive and multiplicative Holt Winter's method*.

An interesting finding regarding estimating the effects of the CPC set in AdWords is that *there is a significant relation between the CPC set and the impression share*. The AdWords model currently used does not take this relation into account. The indirect relation between the CPC and the CTR are also present when grouping keywords based on their destination URL.

Regarding the conversion rate of a product group we found that the progress of the conversion rate is in the same line as the progress of the conversion rate for a whole country. *By levelling the forecast for the conversion rate at a country level, which is based on more data, a forecast can be made for a product group.* This means that the development of the conversion rate at a country level is comparable to the development of the conversion rate for a product group; however the height of the conversion rate differs. Therefore we multiply the forecasted conversion rate at the country level with the ratio between the conversion rate at country level and product group level.

For forecasting the number PAX per booking we found clear patterns in the number of PAX per booking over the time remaining before departure. *By grouping the bookings based on the travel length of the holiday a more accurate forecast for the average number of PAX per booking can be made than by just choosing the average number of PAX per booking*. For example the average number of PAX per bookings for holidays with a duration of 8 or less days the average number of PAX per booking is declining when the number of weeks before departure decreases.

With the results of forecasting analyses in this chapter we know how to forecast the different parameters used in the Integrated Pricing and Advertising model. It is possible to improve the way of forecasting the different parameters by taking more factors into account. For example when forecasting the number of baseline visits the holidays could be taken into account. Improving the forecasts could be considered if the effects of the model are limited due to inaccurate forecasts.

In the next chapter we perform a simulation study regarding the performance of the proposed Integrated Pricing and Advertising model.

7 Simulated case study

In Chapter 5 we introduced a new model for setting the optimal price and CPC for multiple product groups and AdWords group. In this chapter we use the Integrated Pricing and Advertising (IPA) model as proposed in Chapter 5 and the current used model in a simulated case study to answer the fifth research question: "*Is the model created by answering question 3 a valid model and are the expected benefits interesting for practical usage?*". Besides we want to test the performance of the two other possible improvements proposed in Section 3.4. The first possible improvement is to use multiple time periods for the parameters and the variables used: instead of using one time period for the remaining booking period we split the remaining booking period in multiple time periods. The second possible improvement we test is to use the weighted price of the solution instead of using the highest price assigned. In the current situation the highest price assigned to a part of the remaining booking period is used, we want to test if it is better to use the weighted price of all the assigned prices.

In Section 7.1 we introduce the simulation environment. In Section 7.2 we review the results regarding usage of the weighted prices versus usage of the highest price, using one or multiple time periods, and the usage of the current model versus the model proposed.

7.1 Simulation environment

In this section we describe the simulation environment used for testing the performance of the different models. We use a test case that is based on real data regarding the bookings of holidays to Turkey. For simulated cases we obtain results for the profit when the prices and CPCs are set based on the outcomes of the different models considered. In Section 7.1.1 we give an overview of the models used in for our simulations. In Section 7.1.2 we give a description of the test case used and Section 7.1.3 describes how we set up the simulation process.

7.1.1 Models used in the test case

As mention in the introduction of this chapter we want to compare the performance of the integrated model as proposed in Chapter 5 with the model currently used by Sundio. Besides we want to investigate if using multiple time periods leads to an improvement and whether it is better to use the highest price suggested or a weighted price as the new price. In the rest of this chapter we use the reference to the models as shown in Table 4.

Table 4: Overview of the different models used in the simulation environment

	Current model highest price (CH)	**Current model weighted price (CW)**	**Integrated Advertising & Pricing model (IPA)**
1 time period (1)	CH-1	CW-1	IPA-1
2 time periods (2)	CH-2	CW-2	IPA-2
Multiple time periods (M)	CH-M	CW-M	IPA-M

Because in our simulation we need to solve each model for each week in the booking period, the total computation time for the simulation becomes very high when we simulate multiple booking periods. Therefore we do not use the IPA for multiple time periods in our simulations. We need to solve the model for every week in the booking period since the variables should be set regularly based on the new input of for example the forecast of the bookings and the remaining capacity. Note that when it turns out that the model with multiple time periods performs better than the models with one or two time periods, such a model can still be used in practice since when it only has to be solved once a day.

7.1.2 Description of the test case

For our simulation we use test cases which are based on real data related to the holiday destination Turkey. For our test cases we consider a country with two regions, two cities per region, and accommodations divided into two accommodation groups per region. With an accommodation group we refer to a group of accommodations that are in the same product group. Figure 22: Structure of the test case used for simulation shows the structure of the country used for our simulation. The structure is a simplified representation of a real country in order to limit the time needed for performing thousands of simulations.

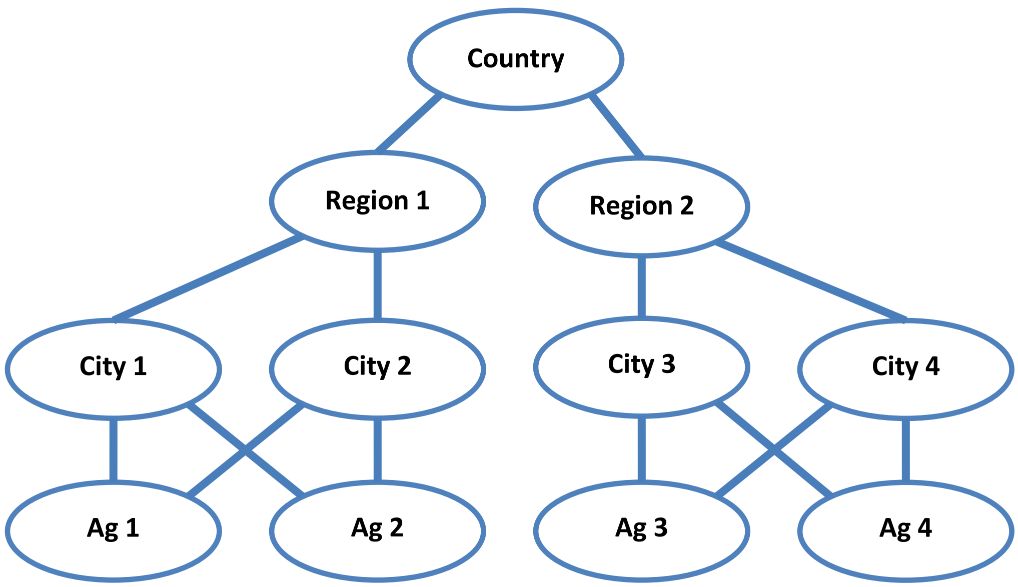

Figure 22: Structure of the test case used for simulation (Ag = Accommodation group)

As mentioned we base the data used for our simulation on real data. Table 8 shows which real data is used to set up the data for the country, regions, cities and accommodations used in our simulation.

Table 5: Overview of data used for the country and the different regions, cities and accommodation groups used in the simulation

In simulation	Data used from
Country	Turkey
Region 1	Turkish Riviera
Region 2	Aegean Coast
City 1	Lara, Alanya
City 2	Antalya, Belek, Kemer, Kundu, Side
City 3	Bitez, Bodrum, Dalyan, Fethiye, Gümbet, Gumulder, Gundogan, Icmeler, Kusadasi
City 4	Marmaris, Oludeniz, Ortakent, Ozdere, Torba, Turgutreis, Türbükü, Turunc,

	Yalikavak
Accommodation group 1	Accommodations in city 1 and 2 sorted in alphabetic order: till "Hotel M"
Accommodation group 2	Accommodations in city 1 and 2 sorted in alphabetic order: starting at "Hotel M"
Accommodation group 3	Accommodations in city 3 and 4 sorted in alphabetic order: till "Hotel M"
Accommodation group 4	Accommodations in city 3 and 4 sorted in alphabetic order: starting at "Hotel M"

The holiday packages used for our simulation have a length of one, two or three weeks. All departures and returns are scheduled after the last day of the week; the first flight is after week 39. The last departure is directly after the last day of week 48. For the simulation data we used the aggregated data of all departures in the week ahead of the departure in real time. This means that for the holiday packages which have a departure directly after week 39 in our simulation, we use the real data of all bookings for holidays with a departure in week 39. We distinguish three travel length groups: short (less than 9 days), medium (9 till 15 days), and long (15 or more weeks). In our simulation these holidays have a length of respectively 1, 2 or three weeks.

Taking into account that we have four accommodation groups, ten departure days and three different lengths for the holiday packages, we consider 120 product groups in our simulation. Besides we have 11 AdWords groups: one at country level, two at region level, four at city level and four at accommodation level. Regarding the resources we consider 20 departure flights (one per region per departure day), 24 return flights (one per region per return day) and 48 capacity restrictions for the accommodation groups (one per week per accommodation group). This means that in total we have 92 resources.

Regarding the costs of the products we assume that the flights costs are fixed costs and the remaining costs of a holiday package are variable costs. This is also the most common situation at Sundio in reality.

7.1.3 Simulation process

In this section we describe how we simulated the part of the booking period considered. We divide our simulation into two parts as illustrated in Figure 23. The first part is simulated based on the real data from the start of week 1, booking year 2011, till the end of week 23 of this booking year. The second part is simulated based on the output of the optimization models and the real data of the period from start week 24, booking year 2013 till the end of week 48 in booking year 2013. We compare the models based on the results in part 2 of the simulation.

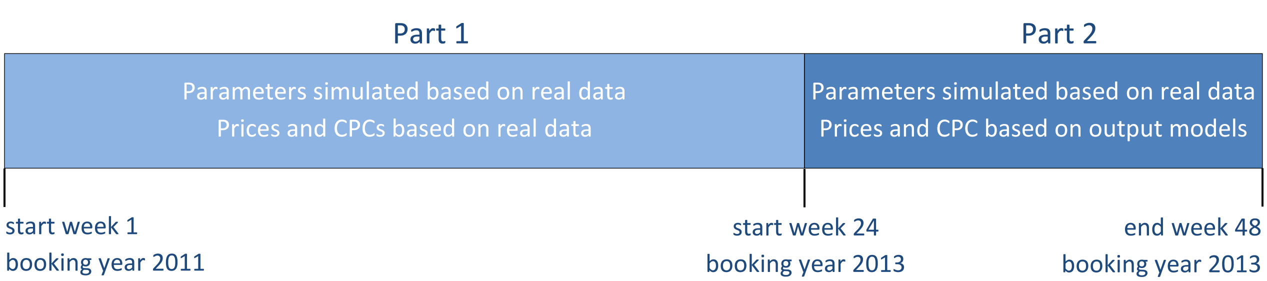

Figure 23: Visualization of different parts of the simulation

The input for the optimization models in based on the data simulated for part 1 of the simulation. For the current model we used the forecasting method as currently used and described in Section 3.1. For the IPA models we use the forecasting methods as proposed in Chapter 6. Regarding the price elasticity and the intercept and slope of the regressions for the AdWords Impression Share we use the simulated values of the last week as a forecast because we are not able to calculate the regressions in AIMMS.

Figure 24 shows the main structure of how a simulation is performed. First we simulate the parameters and variables for the first part of the simulation. Then for every week in part two the model is solved, new prices and CPCs are calculated, and the parameters are simulated. When al weeks are simulated the simulation is done. In the rest of this section we explain how the variables for part 1 and the parameters for the whole simulation are simulated.

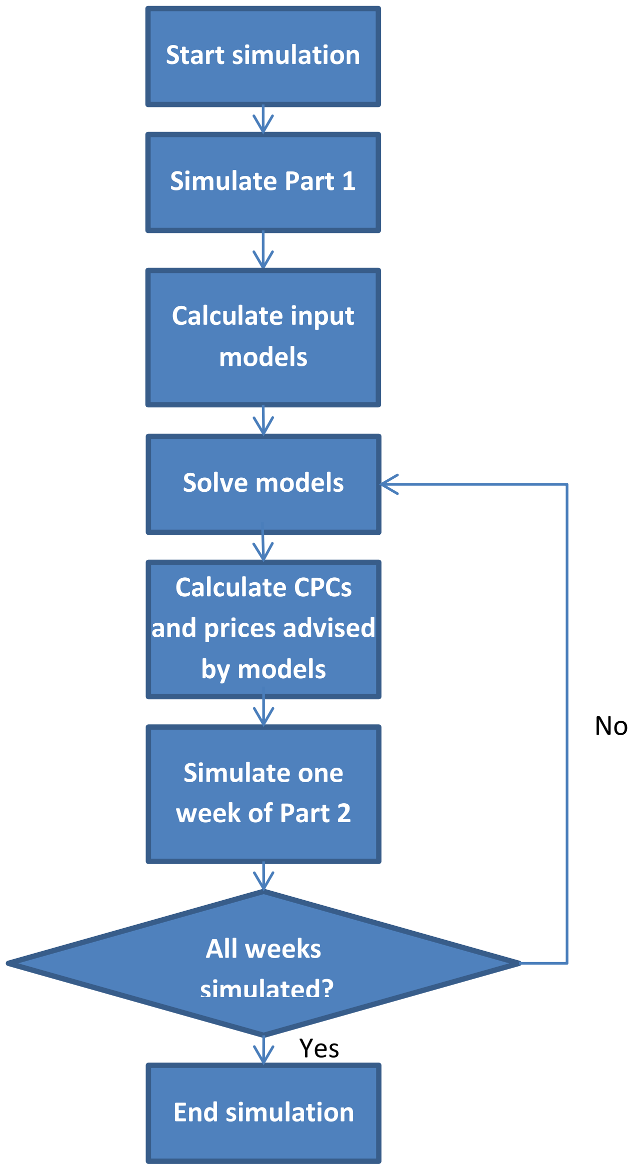

Figure 24: Flow chart of the simulations performed

The parameters which we have to simulate are the number of baseline visits, the number of eligible impressions, the average number of PAX per booking, the price elasticity, the baseline conversion rate, the reference price, the ad ratio and the alphas and betas for the regression of the impression share and the CPC. The prices and CPCs in part 1 are based on real data and in part 2 they are based on the outcome of the models.

To simulate the values of the baseline visits, the number of eligible impressions, the average number of PAX per booking, the price elasticity, the baseline conversion rate and the reference price we first calculate a smoothed average over 5 weeks for the real data of the parameter . We then calculate the mean and standard deviation of the ratio between the smoothed value and the real data points. For every simulated week we take a random pick from a standard normal distribution with this mean and standard deviation; the value picked is used as the ratio between the smoothed value and the simulated value for that week. The price elasticities used as inputs for the simulation are calculated in the same way as that the price elasticities are calculated in the current situation (see Appendix A).

For the ad ratio we use the assumption that the ad ratio is similar to the ratio between the number of baseline visits at a destination page and the product group pages. The simulated ad ratios are calculated based on the numbers of baseline visits simulated.

For the impression share and the click through rate we use the intercepts and slopes calculated from the regressions with the CPC for the different AdWords groups and apply these on the CPC to simulate the impression share and CTR. The mean and the standard deviation of the residuals are used to simulate the variance in the outcomes for the impression share and the CTR over time.

The CPCs and prices used for part one of the simulation are calculated based on real data. We use the same simulation technique as used for simulating the baseline visits, the number of eligible impressions, the average number of PAX per booking, the price elasticity, the baseline conversion rate and the reference price. Based on the number of eligible impressions and by calculating the impression share and click through rate by applying the related regressions on the CPC, we can calculate the number of visits from Google AdWords. By combining the baseline conversion rates, the price elasticities and the prices used we can also calculate the conversion rate.

In the second part of the simulation we constantly solve the model and simulate one extra week based on the output of the models. The different parameters are simulated in the same way as in part 1 of the simulation and based on the output of the models for the prices to use and the CPC to use we calculate the number of clicks from Google AdWords and the number of PAX booked for the different product groups. For the model currently used we calculate the CPC of the next week by using equation 5.10; here the optimal CPC is formulated as $\frac{1}{2}(M - \frac{\alpha}{\beta})$, where M is the expected margin per visitor when using the price proposed by the model and the conversion rate and ad ratio of last week; α and β are parameters describing the relation between the CPC and the CTR ($CTR = \alpha + \beta * CPC$). This is an approximation of the current method used for choosing the CPC used in AdWords; we use this assumption since the current method uses a different definition of AdWords groups as used in our simulation.

The prices used as input for the simulation in Part 2 and the CPC used for the IPA models are based on the output of the different models. Then the impression share and cost per click are calculated based on the simulated alphas and betas for these parameters, and the conversion rate is calculated based on the price elasticity and the base conversion rate simulated in the same way as in part 1.

The number of PAX booked is calculated as the number of visits at the accommodation level of the website times the conversion rate and average PAX per booking of the product group. The capacity is assigned to the product groups based on the index of the product groups. So for example when the demand for the first product group is 3 PAX and for the second product group the demand is 3 PAX, and they both make use of a resource with a capacity of 5 PAX, 3 PAX are assigned to the first product group and 2 PAX are assigned to the second product group.

Based on the numbers of PAX booked for each product group we update the capacities of the resources and again solve the models based on the new input. We perform these calculation for each model separated because when for example the prices proposed by two different models are different, the number of PAX booked will be different and therefore also the input of the models for the next week is different. We repeat this until all weeks till week 48 are simulated. We refer to the simulation of these 25 weeks (week 24 till 48) as the simulation of one booking period.

To reduce the variance between the different models we use the same simulated dataset for all the models. Therefore differences between models are not due to random variance of for example the number of baseline visits.

7.2 Computational results

In this section we compare the performance of the different models when they are applied to the simulated test case as described in the previous section.

To test how the models perform with different remaining capacities we set up three cases. First we performed 250 simulations without taking capacity into account. We do this in order to see what the optimal prices and CPCs are in cases that the capacity is higher than the expected demand. The demand per resource when taking no capacity constraint into account is used to set up three scenarios for the available capacities. For every resource we set the capacity of a resource equal to the lowest, average or maximum demand for that resource for the different simulations where we use no capacity constraint. We use these three scenarios for the capacities as respectively low, average and high capacity available. By doing so we can see how the different models perform for different scenarios of available capacity.

For every capacity scenario we perform 250 simulations. For every simulation we scale the profit of the models to the profit gathered when using the current model of Sundio (CH-1). Figure 25 shows the relative profit of the different models compared to the CH-1 model. We see that the results for the different models based on the current model used are quite similar. Besides we notice that both IPA models perform better than the current model used in most cases. Note that the fixed costs are not taken into account when calculating the relative profits and therefore the relative performance in reality is even better.

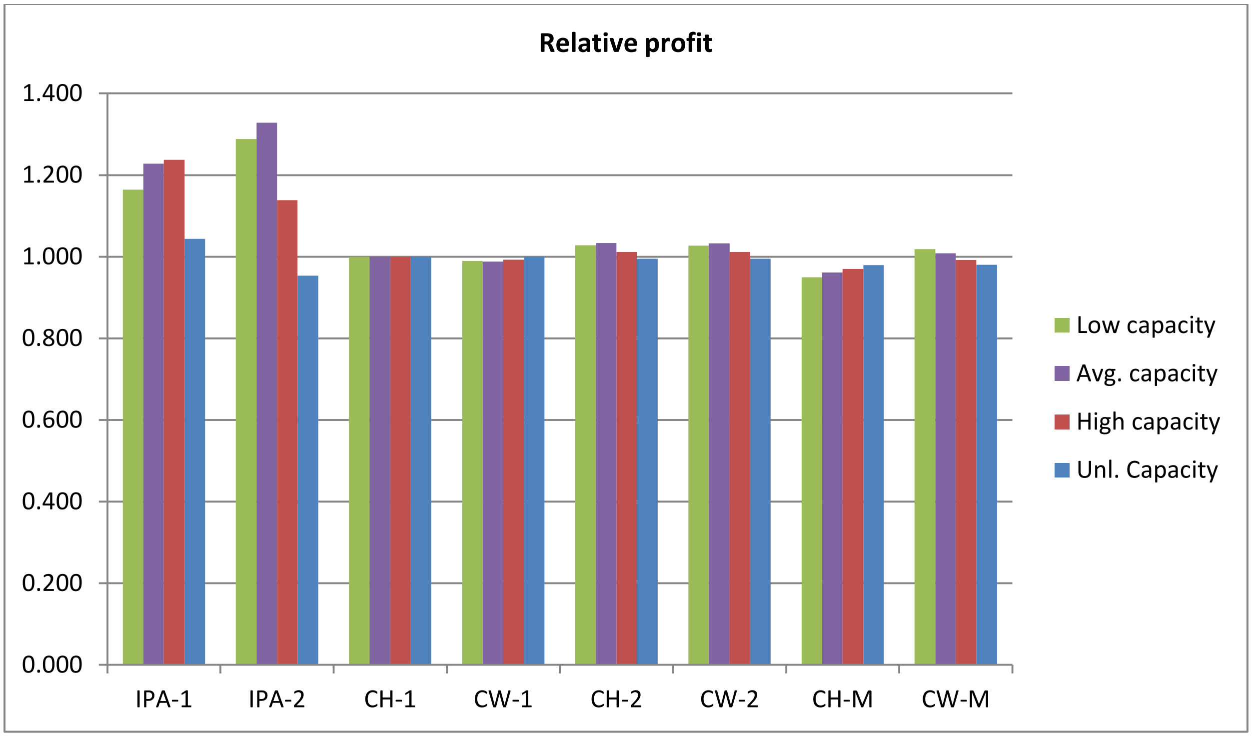

Figure 25: Relative profit of the different models compared to the current model (CH-1) for different capacities

In Table 6 we see the averages and standard deviations of the relative profits retrieved by using the different models in our simulation. The IPA-1 model performs better than the current model, and any other model based on the current model, for any capacity scenario used. The IPA-2 model performs worse than the current model used for the scenario with unlimited capacity, however for the other three scenarios the IPA-2 model performs much better than the IPA-1 model and is able to increase the profit with 13.8 till 28.8% compared to the current model used. Based on the results of this simulated case study we advise to start using the Integrated Pricing and Advertising model as proposed in Chapter 5, however further research is need to see how the models perform in other cases.

Table 6: Relative profit of the different models for different capacities (average (standard deviation))

Capacity	IPA-1	IPA-2	CW-1	CH-2	CW-2	CH-M	CW-M
Low	1.165 (0.055)	1.288 (0.060)	0.990 (0.003)	1.028 (0.009)	1.027 (0.009)	0.950 (0.012)	1.019 (0.011)
Average	1.228 (0.062)	1.328 (0.056)	0.988 (0.004)	1.034 (0.009)	1.033 (0.009)	0.961 (0.010)	1.008 (0.009)
High	1.237 (0.037)	1.138 (0.030)	0.993 (0.004)	1.012 (0.008)	1.011 (0.008)	0.970 (0.009)	0.992 (0.008)
Unlimited	1.044 (0.028)	0.954 (0.013)	1.000 (0.000)	0.995 (0.004)	0.995 (0.004)	0.980 (0.009)	0.980 (0.009)

To investigate why the IPA models outperform the current model we compare the price development and the development of the CPC in Google AdWords. Figure 26 shows the price development for the

IPA-1, IPA-2 and CH-1 model for the scenarios with an average capacity. We see that the price development is quite similar; however the CH-1 model uses a higher price in most weeks although the capacity is not completely used in most cases. The IPA models still have a higher profit since for these models more packages are sold.

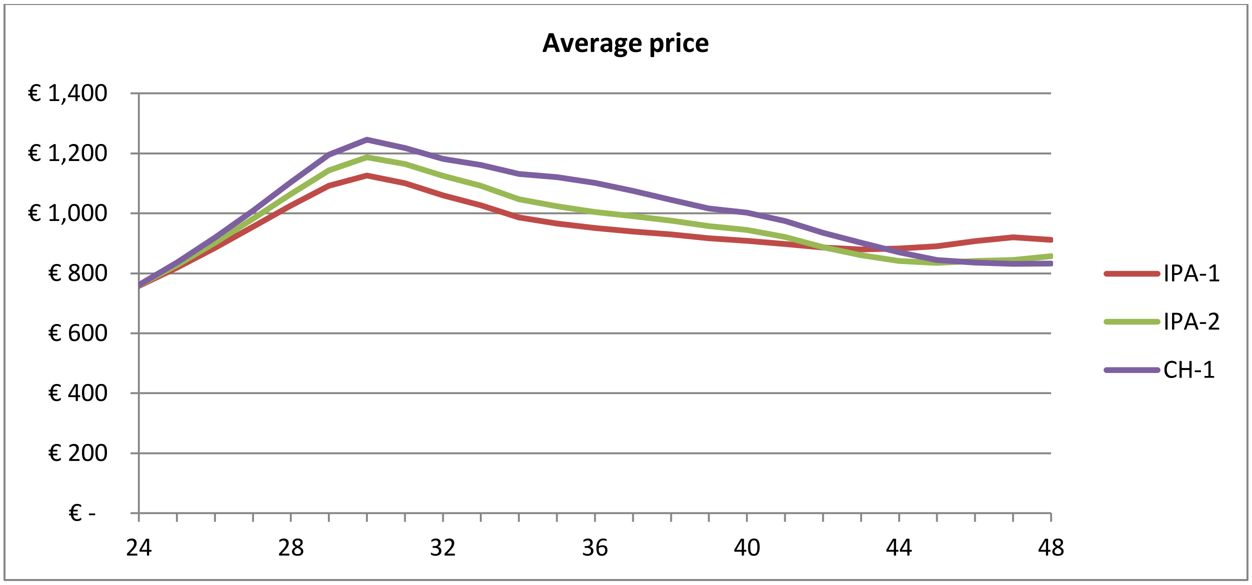

Figure 26: Average price used for the IPA-1, IPA-2 and CH-1 model for the scenario with average capacity

Figure 27 shows the average CPC used for the IPA-1, IPA-2 and CH-1 model for the scenarios with an average capacity. The models based on the current model (CH-2, CH-M, CW-1, CW-2, and CW-3) use a similar CPC as CH-1; we left them out of the graph because they were all overlapping. In this graph we see another reason why the capacity for the CH-1 model is not completely used. The CPC in the CH-1 model is based on the expected revenue of a visitor; however for calculating the optimal CPC the impression share in Google AdWords is not taken into account. When the CPC decreases because of the lower prices used, the impression share will also decrease; this results in a very low impression share and therefore a low amount of visitors and bookings. The API-1 and API-2 model take this effect into account and therefore they keep to CPC stable to attract the needed visitors.

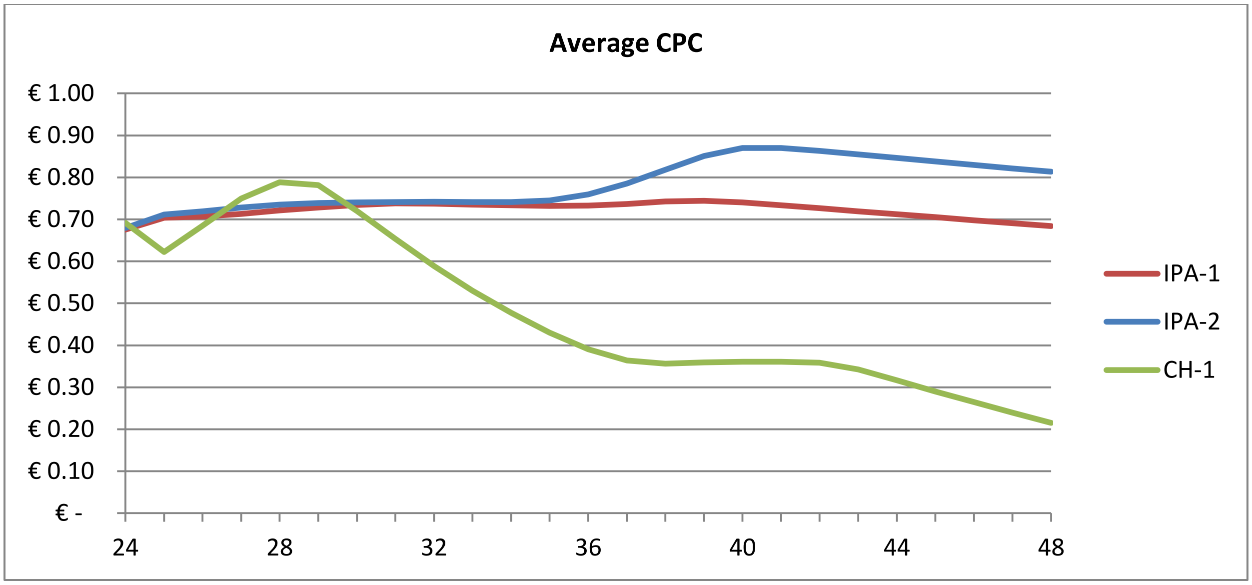

Figure 27: Average CPC used for the IPA-1, IPA-2 and CH-1 model for the scenario with average capacity

We also want to compare how using multiple time periods affects the performance of the model. Therefore we compare the results of the models based on the current model. Table 7 contains the 95% confidence intervals for the relative profit of these models. We see that the CH-M model performs worse than the CH-1 for all different scenarios significantly. The CH-2 model leads to a higher profit than the CH-1 model in all cases except for the unlimited capacity case.

Table 7: 95% confidence intervals for the relative profit of the models compared to the current model

Capacity	IPA-1	IPA-2	CW-1	CH-2	CW-2	CH-M	CW-M
Low	(1.138;1.191)	(1.260;1.317)	(0.988;0.991)	(1.024;1.032)	(1.023;1.031)	(0.944;0.955)	(1.014;1.024)
Average	(1.200;1.257)	(1.300;1.355)	(0.990;0.989)	(1.030;1.038)	(1.030;1.037)	(0.960;0.966)	(1.000;1.013)
High	(1.219;1.255)	(1.124;1.153)	(0.991;0.994)	(1.008;1.015)	(1.007;1.015)	(0.966;0.974)	(0.988;0.996)
Unlimited	(1.030;1.057)	(0.948;0.960)	(1.000;1.000)	(0.993;0.997)	(0.993;0.997)	(0.975;0.984)	(0.975;0.984)

When we compare the models where we used the weighted price we see similar relative performances. The CH-2 and CW-2 models with two time periods perform better than the same model using only one time period (CH-1 and CW-1) with low, average or high capacity. Only for the less realistic case with unlimited capacity the models with one time period perform better. Based on the more realistic capacity cases expanding the current model so that two time periods are used results in an increase in profit (excluding sunk costs) of approximately 1 till 3%. The models with two time periods model outperform the models with multiple time periods in all cases.

For the IPA models we see the same: the model with two time periods outperforms the model with only one time period for the cases with low and average capacity. In the other two capacity scenarios the IPA-1 model performs better, however the case with unlimited capacity is not realistic. Based on these results we advise to start using two time periods for the pricing model, since in most cases all capacity is sold.

Based on the results of the simulations we also want to compare whether using a weighted prices based on the outcomes of the model performs better than using the highest price. In Table 7 we see that the differences between the CH models and the CW models are not significant in most cases. Therefore we see no reason to change the way in which the outcomes of the models for the fractions of price classes to use based on the outcomes of the simulations.

7.3 Conclusions

In this chapter we described a simulated case study that we performed to test the performance of the Integrated Pricing and Advertising model proposed and the model currently used by Sundio. We set up a test case based on real life data related to the destination country Turkey. We used different scenarios for the remaining capacity of the different resources.

The Integrated Pricing and Advertising model outperforms the models which are based on the current model used by Sundio in almost any scenario. Since the IPA-2 model performs better than the IPA-1 model in the most realistic scenarios with low or average capacity compared to the demand *our advice based on this simulated case study is to use the IPA-2 model.*

The main reason for the difference in performance between the Integrated Pricing and Advertising model and the model currently used is that the IPA model takes into account the effects of the CPC on the impression share. We saw that in the simulation performed the CPC was decreased a lot which resulted in a low impression share, while the price was high. The low impression share resulted in less clicks, less visitors attracted and therefore less bookings made.

Based on the outcomes of the simulations we also compared how using 1 time period, 2 time periods or 1 time period for every remaining week influence the performance of the models. We conclude that using 2 time periods results in the best performance. When comparing the results for using the highest price assigned or using a weighted price based on the ratios assigned to the different price classes we did not see a (significant) increase in performance.

8 Conclusions & discussion

In this chapter we answer the main research question:

"What is an appropriate model for Sundio to support the decision making on both prices and advertisements in order to maximize profit?"

We answer this question by giving a conclusion of this research in Section 8.1. This conclusion also contains the recommendations based on this research. In Section 8.3 we discuss the limitations of this research. We mention some opportunities for further research in Section 8.4.

8.1 Conclusions

In this section we answer the main research question of the research performed. Besides we come with some recommendations for Sundio to improve the decision making on both prices and advertisements. The main research question of this research is:

"What is an appropriate model for Sundio to support the decision making on both prices and advertisements in order to maximize profit?"

In Chapter 3 we described the current situation regarding the pricing and advertising models used at Sundio. Sundio uses a pricing model to optimize the prices of all the different packages offered to customers at the website. The input for this model is based on an enhanced forecasting method to build a demand curve indicating the expected demand for using different prices. A model taking into account the expected demand and the capacity restrictions is used to optimize the prices set for the different holidays.

The most important model used for optimizing advertisement decisions is the model used for setting the bids for the advertisements of Sundio in Google AdWords. The Google AdWords model used takes into account the effects between the CPC, the position of the advertisement, the click through rate related to the position, the number of conversion expected based on the CTR and the expected average margin per conversion.

In Chapter 3 we also discussed some opportunities for improvements. We expected that integrating both the results of changing prices and advertisement expenditures can result in finding new opportunities for increasing profit and take away the counteracting effect of both models. Besides we mentioned that the model could be improved by using multiple time periods and by using the weighted price based on the ratios assigned to the different price classes.

We performed a literature study in Chapter 4 to get a better insight in Revenue Management. At Sundio the four steps in the Revenue Management can be distinguished as: (1) historical data is gathered from the data warehouse, (2) a forecast for the demand function is made, (3) optimal price allocations are determined by using a mathematical model, and (4) the optimal prices found are offered to customers via the website. The new model which we propose in the next section influence step 3, the allocation of optimal prices, since the advertisement decisions are integrated in the model. Therefore also step 2, making a forecast for the demand function, changes since some extra forecasts have to be made.

Besides in Chapter 4 we discussed different forecasting methods. We made a distinction between time series forecasting methods and casual forecasting methods. The forecasting methods discussed are taken into account when analysing how to forecast parameters of the proposed model which are currently not forecasted.

In Chapter 5 we propose a new model which integrates the pricing and advertising decisions made by Sundio. We first introduced a simple model which can be used for optimizing cases with just one product group and one AdWords group without taking into account the capacity restrictions. We extend this model by taking into account the capacity restrictions, multiple AdWords groups, and multiple product groups. Besides we constructed the model in such a way that different prices and CPCs can be assigned over time.

The main benefit of the Integrated Pricing and Advertising model proposed compared to the current model used is that it combines the effect of changing the price and changing the CPC in one model. Besides the optimization of the CPCs set in AdWords improves by taking into account the capacity of a product. By using multiple time periods the model also takes expected future price changes into account.

The new model proposed needs extra input compared to the current model used. In Chapter 6 we analyse which forecasting methods should be used for forecasting the different parameters used in the Integrated Pricing and Advertising model. An interesting finding regarding forecasting the effects of the CPC set in AdWords is *that there is a significant relation between the CPC set and the impression share*. By using a higher CPC the chance that the advertisement is show to the person searching for the related keywords increases. The AdWords model currently used does not take this relation into account.

To test the performance of the Integrated Pricing and Advertising model proposed we performed a simulated case study. We performed simulation for different scenarios based on the real life data related to the destination country Turkey. The main conclusion based on the simulated case study is that *the Integrated Pricing and Advertising model performs better than the current model used* in all cases, except for the case with unlimited capacity. Besides we concluded that using one small and one larges time period instead of only one time period also results in an average performance increase of 1 till 3%.

8.2 Recommendations

Based on the results of this study we come up with the following recommendations:

- Start with changing the current model used into a model using one small time period for the close future and one large time period for the remaining time. This change does not need a large investment but in the simulation study performed this change resulted in an increase in profit of 1 till 3% depending on the capacity available.

- Investigate how the effects of the CPC on the impression share can be taken into account in the current model used for Google AdWords. Not taking the effects into account could result in choosing a low CPC because the number of impressions is forecasted to high.
- Further analyse the possible effects of using the Integrated Pricing and Advertising model. The results of the simulated case study performed are interesting, however this is only based on one case study and therefore further research is needed. In the case study performed improvements of almost 30% in profit were noticed for a case with low capacity. For the case with high capacity an improvement in profit of almost 15% was achieved.

8.3 Discussion

In the previous section we answered the main research question of this research. In this section we discuss some limitations regarding the research performed.

The first limitation of this research is related to the conversion rate of the visitors retrieved via Google AdWords. In this research we assumed that the conversion rate for these visitors is the same as for the visitors attracted via other sources. Besides we assumed that the conversion rate is the same for customers attracted via different advertisement. This is partly compensated by using the ad ratio for different destination URLs, because this is used to estimate what percentage of the people visiting the destination URL will also visit an accommodation page; however this ad ratio is only based on the visiting behaviour of the baseline visits. When more information is available about the conversion rates for people attracted via Google AdWords, this information can easily be incorporated in the model proposed. Now both the number of baseline visits and the number of visitors from AdWords are multiplied with the same conversion rate; one could add an extra parameter for the conversion rate for (different) AdWords Groups and multiply the visitors from these AdWords Groups with this conversion rate.

Another limitation of the research performed is also related to Google AdWords. The difficulty regarding the number of visitors from Google AdWords is that it is not know that the visitors attracted would not have visited the website if the advertisement would not have been shown to them. For example when someone searches for the name of an apartment in the Turkish Riviera it is possible that he visits the website by clicking on an organic search result for Sunweb when the advertisement was not shown.

There are also some limitations regarding the forecasting methods advised for forecasting different parameters. The analyses performed regarding the forecasting are limited to the destination country Turkey. It is possible that the advised forecasting methods are not optimal for other destination countries. When Sundio considers using the proposed model we advise to further analyse which methods should be used for forecasting the different parameters.

For the simulated case study performed we only used the data of the destination country Turkey. All though this is the most popular destination country based on the number of PAX booked we doubt whether the results of the simulated case study are enough to draw conclusion for the effects of the IPA model for all destinations of Sundio.

8.4 Further research

In the previous section we discussed the limitations of our research. In this section we mention some possibilities for further research based on these limitations and on the experiences gained during this research. As mentioned in the previous section it is difficult to draw solid conclusion regarding the performance of the IPA model based on the simulated case study performed only. Therefore further research is needed to investigate the effects of the IPA model for the complete selection of holidays offered by Sundio.

For the Integrated Pricing and Advertising model we focused on the advertisement decisions related to Google AdWords. We think that for further research it would be interesting to investigate the possibilities of integrating other advertising methods into the pricing model. For most online advertising platforms it is possible to integrate them into the current model in a similar way as the integration of Google AdWords; however for offline advertising it is more difficult to expand the model. One of the difficulties would be that advertisements on most offline advertising platforms have to be planned further in advance. Regarding Google AdWords one could choose to start or remove an advertisement at any given moment, while air time for TV commercials for example has to be bought months in advance.

Another interesting possibility for further research is related to the upcoming trend of dynamic packaging for holidays. This means that customers do not have to buy a predefined package of transport and accommodation, but they can build their own package. When this method of offering packages becomes widely used, the pricing of the different travel modules becomes more important, especially when customers are able to compare the offers of different suppliers. Investigating how the model should be changed to optimize pricing and advertising decisions in a dynamic packaging context is an interesting opportunity for further research.

In this research we focused on setting up a model for the online tour operator Sundio; it is also interesting to investigate in which other sectors a similar model could be used. We think that the model can also be used at other companies which have to deal with revenue management and advertising. For example a hotel could use a similar model for the pricing of their different rooms and by segmenting AdWords groups based on the type of rooms booked by the people coming from those advertisements. Another example is the car rental industry; one could consider to segment AdWords groups based on the geographical location of the people using Google. For example when many cars are available in one region one could choose to target ads to related geographic locations.

References

De Vries, G., & Pak, K. (2011). Conquering the one-size-fits-all disease: How to put revenue management into action. *Journal of Revenue & Pricing Management, 10*(6), 484-491.

DeLurgio, S. A. (1998). *Forecasting Principles and Applications*: Irwin/McGraw-Hill.

Gardner, E. S., & Dannenbring, D. G. (1980). Forecasting with exponential smoothing: Some guidelines for model selection. *Decision Sciences, 11*(2), 370-383.

Gardner Jr, E. S., & McKenzie, E. (1985). Forecasting trends in time series. *Management Science, 31*(10), 1237-1246.

Google. (2000). Google launches self-service advertising program. Retrieved September 23, 2013, from http://googlepress.blogspot.nl/2000/10/google-launches-self-service.html

Google. (2013). Check and understand Quality Score. Retrieved December 10, 2013, from http://googlepress.blogspot.nl/2000/10/google-launches-self-service.html

Hanke, J. E., & Reitsch, A. G. (1998). *Business Forecasting*: Prentice-Hall International.

Hanssens, D. M., Parsons, L. J., & Schultz, R. L. (2003). *Market response models: Econometric and time series analysis* (Vol. 12): Springer.

Hopp, W. J., & Spearman, M. L. (2008). *Factory Physics*: McGraw-Hill Irwin Irwin.

Kimes, S. E. (1989). Yield management: a tool for capacity-considered service firms. *Journal of operations management, 8*(4), 348-363.

McGill, J. I., & Van Ryzin, G. J. (1999). Revenue management: Research overview and prospects. *Transportation science, 33*(2), 233-256.

ORTEC. (2013). Company profile. Retrieved September 23, 2013, from http://www.ortec.com/about/company_profile.aspx

Segura, J. V., & Vercher, E. (2001). A spreadsheet modeling approach to the Holt–Winters optimal forecasting. *European Journal of Operational Research, 131*(2), 375-388.

Smith, B. C., Leimkuhler, J. F., & Darrow, R. M. (1992). Yield management at American airlines. *Interfaces, 22*(1), 8-31.

Sundio Group. (2013a). De organisatie. Retrieved September 23, 2013, from http://www.sundiogroup.com/bedrijf.html

Sundio Group. (2013b). Nieuws. Retrieved September 10, 2013, from http://www.sundiogroup.com/nieuws.html

Talluri, K. T., & Van Ryzin, G. J. (2005). *The theory and practice of revenue management* (Vol. 68): Springer.

Weatherford, L. R., & Bodily, S. E. (1992). A taxonomy and research overview of perishable-asset revenue management: yield management, overbooking, and pricing. *Operations Research, 40*(5), 831-844.

Winston, W. L. (2004). *Operations research: applications and algorithms*: Brooks/Cole - Thomson Learning.

Yaffee, R. A., & McGee, M. (2000). *Introduction to Time Series Analysis and Forecasting: With Applications of SAS and SPSS*: Academic Press.

Appendix B. Bookings Turkish Riviera example product

Table 9: Bookings Turkish Riviera example product Section 3.1

Weeks Before Departure	Cum. PAX 2011	Cum. PAX 2012	Percentage 2011	Percentage 2012	Average Percentage
0	144	163	100.0%	100.0%	100.0%
1	131	145	91.0%	89.3%	90.1%
2	113	132	78.3%	81.3%	79.8%
3	94	122	65.4%	75.0%	70.2%
4	76	116	52.5%	71.3%	61.9%
5	66	104	45.6%	64.2%	54.9%
6	61	94	42.0%	57.9%	50.0%
7	59	88	40.7%	54.4%	47.5%
8	55	80	37.8%	49.2%	43.5%
9	50	74	34.3%	45.3%	39.8%
10	44	71	30.6%	43.4%	37.0%
11	41	69	28.7%	42.4%	35.6%
12	38	67	26.5%	41.0%	33.7%
13	36	62	25.1%	38.3%	31.7%
14	34	58	23.5%	35.5%	29.5%
15	33	53	23.0%	32.8%	27.9%
16	32	50	22.0%	30.6%	26.3%
17	31	49	21.3%	30.1%	25.7%
18	29	48	20.2%	29.3%	24.7%
19	29	45	20.2%	27.7%	23.9%
20	27	42	18.4%	26.1%	22.2%
21	26	41	18.1%	25.5%	21.8%
22	26	41	18.1%	25.5%	21.8%
23	26	39	17.9%	23.7%	20.8%
24	25	39	17.2%	23.7%	20.5%
25	24	37	16.9%	22.6%	19.8%
26	22	37	15.1%	22.6%	18.9%
27	18	34	12.7%	21.2%	16.9%
28	18	34	12.7%	21.2%	16.9%
29	16	31	11.1%	18.9%	15.0%
30	14	29	9.6%	18.1%	13.8%
31	12	25	8.2%	15.6%	11.9%
32	10	22	6.9%	13.5%	10.2%
33	8	17	5.8%	10.5%	8.1%
34	8	14	5.8%	8.7%	7.3%
35	8	14	5.8%	8.4%	7.1%
36	7	11	5.0%	6.6%	5.8%
37	6	10	4.3%	6.2%	5.2%
38	5	6	3.7%	3.8%	3.8%
39	4	5	2.8%	3.2%	3.0%
>=40	4	4	2.8%	2.5%	2.6%

Appendix C. Finding the optimal price for the model described in Section 5.1

We have to maximize

$$Z = (P-q)*(b+\alpha+\beta*CPC)*\left(cr*\left(1+e*\frac{P-rp}{rp}\right)\right)*pb - CPC*(\alpha+\beta*CPC)$$

Using $V=(b+\alpha+\beta*CPC)$ and $C=CPC*(\alpha+\beta*CPC)$ we can write

$$Z = V*cr*pb*P+\frac{V*cr*e*pb*P^2}{rp}-V*cr*e*pb*P-V*cr*pb*q$$
$$+\frac{V*cr*e*pb*q*P}{rp}+cr*e*pb*q-C$$

We set the derivative over P equal to zero to find the optimal P noted as P^*.

$$\frac{dZ}{dP}=V*cr+\frac{2V*cr*e*pb*P^*}{rp}-V*cr*e*pb-\frac{V*cr*e*pb*q}{rp}=0$$

$$1+\frac{2e*P^*}{rp}-e-\frac{e*q}{rp}=0$$

$$rp+2*e*P^*-e*rp-e*q=0$$

$$\frac{rp}{e}+2P^*-rp-q=0$$

$$2P^*=rp+q-\frac{rp}{e}$$
$$P^*=\frac{1}{2}(rp+q-\frac{rp}{e})$$

Appendix D. Finding the optimal CPC for the model described in Section 5.1

We have to maximize

$$Z = (P^{*} - q) * (\mathrm{bv} + \alpha + \beta * \mathrm{CPC}) * \left(\mathrm{cr} * \left(1 + \mathrm{e} * \frac{\mathrm{P}^{*} - \mathrm{rp}}{\mathrm{rp}}\right)\right) * pb - CPC * (\alpha + \beta * CPC)$$

We can rewrite this as

$$Z = M * (bv + \alpha + \beta * CPC) - CPC * (\alpha * \beta * CPC)$$

where

$$M = (P^{*} - q) * (cr * \left(1 + e * \frac{(P^{*} - rp)}{rp}\right) * pb$$

We set the derivative over CPC equal to zero to find the optimal CPC noted as CPC^{*}.

$$\frac{dZ}{dCPC} = M * \beta - \alpha - 2\beta * CPC^{*} = 0$$

$$2\beta * CPC^{*} = M * \beta - \alpha$$

$$CPC^{*} = \frac{1}{2}(M - \frac{\alpha}{\beta})$$

Appendix E. Text representation AIMMS model with one product and one AdWords group

```
SET:
  identifier :  PriceClass
  index      :  j ;

SET:
  identifier :  CPCClass
  index      :  h

PARAMETER:
  identifier :  VarCost ;

PARAMETER:
  identifier   :  Price
  index domain :  (j) ;

PARAMETER:
  identifier   :  CPC
  index domain :  (h) ;

PARAMETER:
  identifier   :  BaseVisits ;

PARAMETER:
  identifier   :  AdVisits
  index domain :  (h) ;

PARAMETER:
  identifier   :  ConvRate
  index domain :  (j) ;

PARAMETER:
  identifier   :  PaxBooking ;

PARAMETER:
  identifier   :  Cap ;

VARIABLE:
  identifier   :  Fraction
```

```
  index domain :  (j,h)
  range       :  [0, 1] ;

 VARIABLE:
  identifier  :  Profit
  range       :  free
  definition      :     sum[j,sum[h,((Price(j)-VarCost)*(BaseVisits+AdVisits(h))*ConvRate(j)*PaxBooking-
AdVisits(h)*CPC(h))*Fraction(j, h)]] ;

 PARAMETER:
  identifier  :  Copy_Profit
  range       :  free
  definition  :  sum[j,sum[h,((Price(j)-VarCost)*(BaseVisits+AdVisits(h))*ConvRate(j)*PaxBooking)]] ;

 MATHEMATICAL PROGRAM:
  identifier  :  MaxProfit
  objective   :  Profit
  direction   :  maximize
  constraints :  Con
  variables   :  Var
  type        :  Automatic ;

 CONSTRAINT:
  identifier  :  CapLim
  definition  :  sum[j,sum[h,((BaseVisits+AdVisits(h))*ConvRate(j)*PaxBooking)*Fraction(j, h)]]<=Cap ;

 CONSTRAINT:
  identifier  :  SumFrac
  definition  :  sum[j,sum[h,Fraction(j, h)]]<=1 ;
```

Appendix F. Text representation AIMMS model with one product and one AdWords group with time element

```
SET:
   identifier :  PriceClass
   index      :  j ;

SET:
   identifier :  CPCClass
   index      :  h

SET:
   identifier :  Time
   index      :  t ;

 PARAMETER:
   identifier :  VarCost ;

 PARAMETER:
   identifier   :  Price
   index domain :  (t,j) ;

 PARAMETER:
   identifier   :  CPC
   index domain :  (h) ;

 PARAMETER:
   identifier   :  BaseVisits
   index domain :  (t) ;

 PARAMETER:
   identifier   :  AdVisits
   index domain :  (t,h) ;

 PARAMETER:
   identifier   :  ConvRate
   index domain :  (t,j) ;

 PARAMETER:
   identifier   :  PaxBooking
   index domain :  (t) ;

 PARAMETER:
   identifier   :  Cap ;

 VARIABLE:
   identifier   :  Fraction
   index domain :  (t,j,h)
   range        :  [0, 1] ;

 VARIABLE:
```

```
    identifier  : Profit
    range       : free
    definition  : sum[t,sum[j,sum[h,((Price(t,j)-VarCost)*(BaseVisits(t)+AdVisits(t,h))*ConvRate(t,j)*PaxBooking(t)-
AdVisits(t,h)*CPC(h))*Fraction(t, j, h)]]] ;

  MATHEMATICAL PROGRAM:
    identifier  : MaxProfit
    objective   : Profit
    direction   : maxize
    constraints : Con
    variables   : Var
    type        : Automatic ;

  CONSTRAINT:
    identifier  : CapL
    definition  : sum[t,sum[j,sum[h,((BaseVisits(t)+AdVisits(t,h))*ConvRate(t,j)*PaxBooking(t))*Fraction(t, j,
h)]]]<=Cap ;

  CONSTRAINT:
    identifier  : SumFrac
    index domain : (t)
    definition  : sum[j,sum[h,Fraction(t, j, h)]]<=1 ;
```

Appendix G. Text representation AIMMS model with multiple product and one AdWords groups with time element

```
SET:
  identifier  : Time
  index       : t ;

SET:
  identifier  : ProductGroups
  indices     : i ;

SET:
  identifier  : AdWordsGroups
  index       : g ;

SET:
  identifier  : Resources
  index       : k ;

SET:
  identifier  : PriceClass
  index       : j

SET:
  identifier  : CPCClass
  index       : h

PARAMETER:
  identifier  : VarCost
  index domain :  (i) ;

PARAMETER:
  identifier  : ProductResource
  index domain :  (i,k) ;

PARAMETER:
  identifier  : AdWordsProduct
  index domain :  (i,g) ;

PARAMETER:
  identifier  : RelDur
  index domain :  (t,i) ;
```

```
PARAMETER:
  identifier   : AdRatio
  index domain : (t,i,g) ;

PARAMETER:
  identifier   : Price
  index domain : (t,i,j) ;

PARAMETER:
  identifier   : Capacity
  index domain : (k) ;

PARAMETER:
  identifier   : CPC
  index domain : (g,h) ;

PARAMETER:
  identifier   : BaseVisits
  index domain : (t,i) ;

PARAMETER:
  identifier   : AdVisits
  index domain : (t,g,h) ;

PARAMETER:
  identifier   : ConvRate
  index domain : (t,i,j) ;

PARAMETER:
  identifier   : PaxBooking
  index domain : (t,i) ;

PARAMETER:
  identifier   : Cap
  index domain : (k) ;

VARIABLE:
  identifier   : Fraction
  index domain : (t,i,j,g,h)
  range        : [0, 1] ;

VARIABLE:
  identifier   : FractPrice
  index domain : (t,i,j)
  range        : free ;
```

```
VARIABLE:
  identifier  : FractCPC
  index domain : (t,g,h)
  range      : free ;

VARIABLE:
  identifier  : Profit
  range      : free
  definition  : sum[t,sum[(i,j),((Price(t,i,j)-
VarCost(i))*RelDur(t,i)*(FractPrice(t,i,j)*BaseVisits(t,i)+sum[(g,h)|AdWordsProduct(i,g),AdRatio(t, i,
g)*AdVisits(t,g,h)*Fraction(t,i,j,g,h)])*ConvRate(t,i,j)*PaxBooking(t,i))]-
sum[(g,h),AdVisits(t,g,h)*CPC(g,h)*FractCPC(t,g,h)]] ;

MATHEMATICAL PROGRAM:
  identifier  : MaxProfit
  objective   : Profit
  direction   : maxize
  constraints : Con
  variables   : Var
  type       : Automatic ;

CONSTRAINT:
  identifier  : CapL
  index domain : k
  definition  :
sum[t,sum[(i,j),RelDur(t,i)*(FractPrice(t,i,j)*BaseVisits(t,i)+sum[(g,h)|AdWordsProduct(i,g),AdRatio(t, i,
g)*AdVisits(t,g,h)*Fraction(t,i,j,g,h)])*ConvRate(t,i,j)*PaxBooking(t,i)]]<=Cap(k) ;

CONSTRAINT:
  identifier  : OnePrice
  index domain : (t,i)
  definition  : sum[j,FractPrice(t,i, j)]<=1 ;

CONSTRAINT:
  identifier  : OneCPC
  index domain : (t,g)
  definition  : sum[h,FractCPC(t,g, h)]<=1 ;

CONSTRAINT:
  identifier  : CombPrice
  index domain : (t,i,j,g)
  definition  : sum[h,Fraction(t,i, j, g, h)]=FractPrice(t,i, j) *RelDurM(t,i) ;

CONSTRAINT:
  identifier  : CombCPC
  index domain : (t,i,g,h)
  definition  : sum[j,Fraction(t,i,j,g,h)]=FractCPC(t,g,h) ;
```

Appendix H. ANOVA tables for the Running Time versus the number of Variables, Constraints and Nonzeros

Regression Statistics	
Multiple R	0.972057
R Square	0.944894
Adjusted R Square	0.944893
Standard Error	0.037808
Observations	200000

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	4902.114	1634.038	1143103	0
Residual	199996	285.8895	0.001429		
Total	199999	5188.004			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.005939	0.000108	54.99062	0	0.005727	0.00615
Variables	-1.4E-05	6.38E-08	-223.987	0	-1.4E-05	-1.4E-05
Constraints	2.87E-05	1.54E-07	186.4516	0	2.84E-05	2.9E-05
NonZeros	2.33E-06	3.26E-09	714.2588	0	2.32E-06	2.34E-06

Appendix I. MAPE values for analyses in Chapter 6

In this appendix the MAPE values for different analyses as discussed in Chapter 6 are shown. The different methods used are:

1. Last week: $\hat{y}_t = y_{t-1}$
2. Same week last year: $\hat{y}_t = y_{t-52}$
3. Average same week last two years: $\hat{y}_t = \frac{y_{t-52}+y_{t-104}}{2}$
4. Calculate ratio between last five weeks this year and previous year and multiply same week last year with this ratio: $\hat{y}_t = \frac{\sum_{i=1}^{5} y_{t-i}}{\sum_{i=1}^{5} y_{t-i-52}} * y_{t-52}$
5. Additive Holt Winters' method with a damped trend and a length for seasonality of 52 week.
6. Multiplicative Holt Winters' method with a damped trend and a length for seasonality of 52 week.

The columns represent different weeks of smoothing. For example in Table 10 using method 2 smoothed over 5 weeks has a MAPE of 0.37 for the product group with 1 accommodation.

Baseline visits

Table 10: Accuracy for forecasting the number of baseline visits in the remaining weeks of the booking period (MAPE)

Method	1 accommodation				11 accommodations				25 accommodations			
	1	3	5	7	1	3	5	7	1	3	5	7
1	0.52				0.45				0.31			
2	0.41	0.38	0.37	0.37	0.36	0.35	0.33	0.32	0.31	0.29	0.28	0.28
3	1.12	1.12	1.11	1.11	0.34	0.32	0.31	0.30	0.21	0.20	0.20	0.19
4	0.83	0.80	0.80	0.78	0.39	0.37	0.35	0.32	0.53	0.51	0.50	0.50
5	1.93	1.92	1.91	1.88	0.34	0.32	0.31	0.30	0.32	0.31	0.30	0.29
6	1.91	1.90	1.90	1.88	0.35	0.32	0.31	0.30	0.56	0.55	0.54	0.52

Table 11: Accuracy for forecasting the number of baseline visits in the next week (MAPE)

Method	1 accommodation				11 accommodations				25 accommodations			
	1	3	5	7	1	3	5	7	1	3	5	7
1	0.15				0.20				0.14			
2	0.37	0.35	0.34	0.33	0.33	0.31	0.30	0.29	0.31	0.28	0.27	0.27
3	0.81	0.80	0.80	0.79	0.31	0.29	0.28	0.28	0.25	0.24	0.24	0.22
4	0.28	0.21	0.21	0.24	0.29	0.25	0.23	0.25	0.26	0.19	0.19	0.21
5	0.34	0.32	0.35	0.49	0.23	0.21	0.20	0.21	0.18	0.16	0.17	0.19
6	0.22	0.18	0.19	0.25	0.26	0.21	0.20	0.22	0.22	0.20	0.19	0.24

Table 12: Accuracy for forecasting the number of baseline visits in the remaining booking period (MAPE)

Method	1 accommodation				11 accommodations				25 accommodations			
	1	3	5	7	1	3	5	7	1	3	5	7
1	0.40				0.50				0.32			
2	0.30	0.30	0.30	0.29	0.26	0.26	0.26	0.26	0.20	0.19	0.19	0.19
3	0.66	0.66	0.66	0.65	0.30	0.30	0.30	0.30	0.18	0.18	0.18	0.18
4	0.54	0.53	0.53	0.52	0.28	0.28	0.27	0.27	0.40	0.40	0.39	0.39
5	0.92	0.91	0.91	0.90	0.28	0.28	0.28	0.27	0.23	0.22	0.22	0.22
6	0.79	0.79	0.79	0.79	0.28	0.28	0.27	0.27	0.34	0.33	0.33	0.32

Eligible impressions

Table 13: Accuracy for forecasting the number of eligible impressions in Google AdWords in the remaining weeks of the booking period (MAPE)

Method	Turkey				Turkish Riviera				Alanya				Accommodations			
	1	3	5	7	1	3	5	7	1	3	5	7	1	3	5	7
1	0.50				2.21				0.49				1.56			
2	0.73	0.65	0.59	0.57	2.82	2.85	2.91	2.97	0.33	0.27	0.25	0.25	4.85	4.85	4.85	4.85
3	1.32	1.27	1.22	1.19	2.22	2.25	2.29	2.33	0.79	0.78	0.79	0.80	3.50	3.51	3.51	3.50
4	0.93	0.86	0.79	0.74	1.97	1.97	1.99	2.04	2.75	2.74	2.77	2.84	4.06	4.06	4.06	4.06
5	3.34	3.28	3.22	3.14	3.04	3.06	3.05	3.04	2.22	2.19	2.18	2.13	1.98	1.92	1.88	1.85
6	5.17	5.06	4.98	4.86	5.18	5.19	5.16	5.12	4.56	4.49	4.43	4.47	1.39	1.39	1.40	1.42

Table 14: Accuracy for forecasting the number of eligible impressions in Google AdWords in the next week (MAPE)

Method	Turkey				Turkish Riviera				Alanya				Accommodations			
	1	3	5	7	1	3	5	7	1	3	5	7	1	3	5	7
1	0.21				0.37				0.16				0.22			
2	0.63	0.56	0.52	0.51	3.17	3.25	3.38	3.52	0.45	0.39	0.35	0.35	3.42	3.42	3.43	3.45
3	1.08	1.06	1.05	1.05	2.27	2.34	2.42	2.50	1.17	1.14	1.15	1.18	2.57	2.57	2.57	2.57
4	0.52	0.39	0.37	0.40	1.05	0.80	0.83	1.00	0.63	0.62	0.72	0.99	0.33	0.27	0.29	0.33
5	0.81	0.80	0.90	1.25	0.80	0.74	0.70	0.82	0.99	0.90	0.92	1.10	0.65	0.56	0.53	0.65
6	1.03	0.90	1.05	1.75	0.92	0.62	0.55	0.68	2.63	3.68	3.32	3.52	0.40	0.35	0.33	0.36

Table 15: Accuracy for forecasting the number of eligible impressions in the remaining booking period (MAPE)

Method	Country				Region				City				Accommodation			
	1	3	5	7	1	3	5	7	1	3	5	7	1	3	5	7
1	0.34				0.83				0.75				0.52			
2	0.25	0.25	0.24	0.24	1.14	1.16	1.18	1.22	0.17	0.17	0.16	0.16	2.93	2.93	2.93	2.92
3	0.71	0.70	0.69	0.68	0.82	0.83	0.84	0.87	0.78	0.78	0.78	0.78	2.21	2.21	2.21	2.21
4	0.47	0.46	0.45	0.43	1.25	1.26	1.25	1.26	2.76	2.75	2.74	2.75	1.57	1.57	1.57	1.57
5	2.47	2.44	2.38	2.27	1.24	1.24	1.23	1.21	1.96	1.94	1.91	1.84	1.24	1.22	1.18	1.11
6	3.94	3.92	3.87	3.75	1.98	1.98	1.95	1.91	6.76	6.42	6.15	5.74	0.87	0.86	0.84	0.81

Ad ratio

Table 16: Accuracy for forecasting the ratio between the numbers of visitors at the destination URL that visits the accommodation page (MAPE)

Method	Turkey				Turkish Riviera				Alanya			
	1	3	5	7	1	3	5	7	1	3	5	7
1	0.57				0.52				0.18			
2	0.41	0.38	0.37	0.37	0.41	0.38	0.37	0.37	0.41	0.38	0.37	0.37
3	1.13	1.12	1.11	1.11	1.13	1.12	1.11	1.11	1.13	1.12	1.11	1.11
4	0.37	0.35	0.35	0.38	0.37	0.35	0.36	0.38	0.20	0.19	0.18	0.18
5	0.69	0.67	0.66	0.66	0.67	0.65	0.65	0.65	0.52	0.51	0.51	0.50
6	0.75	0.73	0.73	0.74	0.72	0.71	0.71	0.72	0.41	0.41	0.40	0.40

Table 17: Accuracy for forecasting the ratio between the numbers of visitors at the destination URL that visits the accommodation page in the next week (MAPE)

Method	Turkey				Turkish Riviera				Alanya			
	1	3	5	7	1	3	5	7	1	3	5	7
1	0.24				0.22				0.08			
2	0.37	0.35	0.34	0.33	0.37	0.35	0.34	0.33	0.37	0.35	0.34	0.33
3	0.81	0.81	0.80	0.79	0.81	0.81	0.80	0.79	0.81	0.81	0.80	0.79
4	0.33	0.26	0.26	0.27	0.31	0.24	0.24	0.26	0.16	0.12	0.11	0.11
5	0.38	0.32	0.32	0.40	0.38	0.31	0.31	0.39	0.17	0.14	0.13	0.15
6	0.32	0.26	0.28	0.37	0.29	0.25	0.27	0.36	0.12	0.11	0.10	0.11

Table 18: Accuracy for forecasting the ratio between the numbers of visitors at the destination URL that visits the accommodation page in the remaining booking period (MAPE)

Method	Turkey				Turkish Riviera				Alanya			
	1	3	5	7	1	3	5	7	1	3	5	7
1	0.49				0.46				0.31			
2	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30
3	0.66	0.66	0.66	0.65	0.66	0.66	0.66	0.65	0.66	0.66	0.66	0.65
4	0.27	0.27	0.26	0.26	0.28	0.28	0.27	0.27	0.13	0.13	0.13	0.12
5	0.34	0.33	0.33	0.32	0.34	0.34	0.34	0.33	0.35	0.35	0.35	0.35
6	0.43	0.43	0.43	0.42	0.41	0.41	0.40	0.40	0.28	0.28	0.28	0.28

PAX per booking

Table 19: Accuracy for forecasting the average PAX per booking (MAPE)

Method	Turkey			
	1	3	5	7
1	0.27			
2	0.08	0.08	0.08	0.08
3	0.10	0.08	0.09	0.09
4	0.21	0.20	0.19	0.18
5	0.17	0.17	0.16	0.15
6	0.17	0.17	0.16	0.15

Table 20: Accuracy for forecasting the average number of PAX per booking in the next week and in the remaining period (MAPE)

Method	Next week			
	1	3	5	7
1	0.35			
2	0.06	0.06	0.07	0.08
3	0.07	0.07	0.07	0.08
4	0.22	0.22	0.22	0.22
5	0.20	0.20	0.20	0.19
6	0.20	0.21	0.20	0.20

Appendix J. ANOVA tables for the Impression share versus Cost Per Click regressions

The tables in this appendix show the results of the regression analysis for the impression share versus the CPC for different AdGroup levels in Google AdWords.

Country level: Turkey

Regression Statistics	
Multiple R	0.475475
R Square	0.226077
Adjusted R Square	0.209252
Standard Error	0.139819
Observations	48

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*	
Regression	1	0.262695	0.262695	13.43742	0.000637	
Residual	46	0.899277	0.019549			
Total	47	1.161971				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.365342	0.11601	3.149236	0.002874	0.131827	0.598858
CPC	0.740933	0.202126	3.665709	0.000637	0.334076	1.147791

Region level: Turkish Riviera

Regression Statistics	
Multiple R	0.397751
R Square	0.158206
Adjusted R Square	0.139906
Standard Error	0.103311
Observations	48

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*	
Regression	1	0.092272	0.092272	8.645197	0.005117	
Residual	46	0.490969	0.010673			
Total	47	0.583242				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.650423	0.062248	10.44882	9.89E-14	0.525123	0.775723
CPC	0.285442	0.09708	2.940272	0.005117	0.09003	0.480854

City level: Alanya

Regression Statistics	
Multiple R	0.92088
R Square	0.84802
Adjusted R Square	0.844716
Standard Error	0.09163
Observations	48

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	2.155028	2.155028	256.6707	1.92E-20
Residual	46	0.38622	0.008396		
Total	47	2.541247			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-0.23504	0.059993	-3.91783	0.000294	-0.3558	-0.11428
CPC	1.541215	0.0962	16.02095	1.92E-20	1.347574	1.734856

Accommodation level

Regression Statistics	
Multiple R	0.62974
R Square	0.396573
Adjusted R Square	0.383455
Standard Error	0.068059
Observations	48

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.140033	0.140033	30.23124	1.62E-06
Residual	46	0.213075	0.004632		
Total	47	0.353108			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.731207	0.031747	23.03246	6.77E-27	0.667304	0.79511
CPC	0.267122	0.048583	5.498294	1.62E-06	0.16933	0.364914

Appendix K. ANOVA tables for the Click Through Rate versus Cost per Click regressions

The tables in this appendix show the results of the regression analysis for the impression share versus the CPC for different AdGroup levels in Google AdWords.

Country level: Turkey

Regression Statistics	
Multiple R	0.740273
R Square	0.548005
Adjusted R Square	0.529172
Standard Error	0.019744
Observations	26

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.011343	0.011343	29.09789	1.54E-05
Residual	24	0.009356	0.00039		
Total	25	0.020699			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-0.02741	0.020917	-1.31037	0.202473	-0.07058	0.015762
CPC	0.201782	0.037407	5.394246	1.54E-05	0.124578	0.278986

Region level: Turkish Riviera

Regression Statistics	
Multiple R	0.73151
R Square	0.535107
Adjusted R Square	0.515736
Standard Error	0.063442
Observations	26

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.111187	0.111187	27.62474	2.18E-05
Residual	24	0.096598	0.004025		
Total	25	0.207785			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-0.18871	0.061037	-3.09169	0.004986	-0.31468	-0.06273
CPC	0.454367	0.086449	5.255924	2.18E-05	0.275946	0.632788

City level: Alanya

Regression Statistics	
Multiple R	0.778581
R Square	0.606188
Adjusted R Square	0.589779
Standard Error	0.003323
Observations	26

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.000408	0.000408	36.94273	2.81E-06
Residual	24	0.000265	1.1E-05		
Total	25	0.000673			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.005015	0.002973	1.686568	0.104644	-0.00112	0.011151
CPC	0.034265	0.005637	6.078053	2.81E-06	0.022629	0.0459

Accommodation level

Regression Statistics	
Multiple R	0.654915
R Square	0.428914
Adjusted R Square	0.405119
Standard Error	0.028923
Observations	26

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.015078	0.015078	18.02519	0.000283
Residual	24	0.020076	0.000837		
Total	25	0.035155			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.055062	0.018011	3.057126	0.005416	0.017889	0.092234
CPC	0.133279	0.031392	4.245609	0.000283	0.068489	0.19807