VRIJE UNIVERSITEIT AMSTERDAM

THE REAL THEORY AND PRACTICE OF REVENUE MANAGEMENT

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor of Philosophy aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Bètawetenschappen
op maandag 21 september 2020 om 15.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Daniël Sebastiaan Hopman

geboren te Aalsmeer

promotoren:	prof.dr. G.M. Koole
	prof.dr. R.D. van der Mei

THE REAL THEORY AND PRACTICE OF REVENUE MANAGEMENT

DANIEL HOPMAN

2015 - 2020

CONTENTS

FOREWORD

> *"Winning is not a sometime thing; it's an all the time thing. You don't win once in a while; you don't do things right once in a while; you do them right all of the time. Winning is a habit. "*
>
> – 1955 , *Vince Lombardi*

I love winning. I love revenue management (RM). And, I love winning in RM.

And with this thesis, I hope to make you, the reader, fall in love with this part of optimisation, too. The three main characteristics of RM, an perishable asset, limited time, and fixed capacity make the problem so exciting. On the one hand, one has fixed time to sell some form of a perishable asset (seat, hotel room, car rental) and there is no going back fixing a mistake. On the other hand, fixed capacity means you cannot simply add or remove units when required.

However, at the age of 17 I wasn't aware of these details. Rather, I was fascinated by availability - the letters that show how many seats were left for each fare class. These numbers could change at any time. It was a much more dynamic environment than, for example, an online store selling goods. So, at the age my peers were falling in love with girls, I fell in love with RM.

But where do you start? Unfortunately, there aren't books on RM, especially for those with a limited understanding of mathematics. So I started with reading papers and a black board. That black board, years later, still proudly shows equations and examples in my old room at my parents' place.

I'll be the first to admit I am not the best mathematician. In fact, in high school I scored very average. When I joined university, things changed. And though things changed, I wasn't the best in class: ironically, I failed a homework assignment on RM. Initially, I developed an interest in calculus, and optimisation quickly followed. I remember the first time we were introduced to dynamic programming. Solving a problem backwards in time was so counter-intuitive it took a couple of weeks before I finally came to grips with it. For me, it was especially important to gain an understanding, as dynamic programming plays a big role in optimisation in RM.

While optimisation is the topic that excited me most, RM is more than that: the inputs, demand forecasts and pricing estimates, are as important, if not more important. And then, of course, there is the practice of RM. This is something that cannot be taught from books, but is something that you need to experience. And that experience I received by working for two big airlines, as well as for hotel companies. It gave me a lot of insight that put me in the unique position to discuss both the real theory and practice of RM. For me, this research was not about obtaining a doctorate, or getting publications. Rather, it was an opportunity for me to highlight the practice of RM, and how I think things we see in practice can be tackled. For me, personally, the applicability of algorithms, heuristics or methods in practice is just as important as impressive mathematics. This is the very reason I called this thesis the **real** theory and practice of revenue management.

I would like to take this opportunity to thank some of the people around me that helped me make this happen.

Firstly, would like to thank my parents, Tammo and Monique, sisters Merel and Iris and aunt Desiree for their ongoing support. Not just for their support during the past five years of my research, but from the day I was born. I would never have been able to achieve the things that I have without the wonderful upbringing I had. And of course my grandparents, who are smiling down at us from high up in the sky.

Secondly, I would like to thank my supervisors Ger Koole and Rob van der Mei for their advice, thinking and laughs. I would not have been able to complete my research without their valuable input. Sometimes harsh - I remember feeling very dumb after not being able to figure out a very simple dynamic programming formulation, sometimes nice, but this input was always helpful. They also helped me to see the bigger picture. Research is amazing - the ability to transfer new knowledge to someone, but it should also be restricted at some point, they taught me I'm unable to study the wide array of topics I had in mind. Rob van der Mei arranged my internship at KLM Royal Dutch Airlines, and for this I am still extremely grateful since this jumpstarted my professional career.

Lastly, I want to thank my old colleagues-turned-friends Laurens van Venrooy, Antti Tolvanen, Lloyd Misquitta and Manjinder Singh for sharing their knowledge of RM, systems and the practice of RM. They have all played a big role in my professional career. Before I started my professional career, I was book-smart: I knew what forecasting methods were used and how the optimisation algorithms worked. But I realized very quickly that in practice, things do not always work this way, for different reasons. Laurens, Antti, Lloyd and Manjinder helped me understand the intricacies.

Now, read on and be prepared to fall in love with RM.

SUMMARY

This thesis discusses the *real* theory and practice of revenue management (RM). Over the past few decades, there has been extensive research in RM that has led to interesting insights. However, many years of experience in hotel and airline RM has taught that there are still significant gaps between the theory of RM and what is applied in daily practice.

The stage is set in Chapter 1, where gaps between theory and practice in forecasting, pricing, optimisation and culture are addressed. We claim that traditional demand forecasting methods cannot be applied because of a limited number of data points, cancellations are rarely considered explicitly, and the most often used unconstraining method is not reliable. The gaps in pricing we have identified are as follows: the theory is focused on the estimation of price elasticities and willingness to pay, while in practice it is often solely influenced by competition, distrust in the revenue management system and competing KPIs. For optimisation, the theory underestimates the number of products, produces methods that are slow to solve (often a result of large state spaces) and cancellations are not considered to be time-dependent. When culture is considered, the theory assumes perfect information, lack of observation bias and it is assumed customers are "honest", and not using exploits. In summary, it is concluded that the theory does not align with practice since it does not consider (time-dependent) underlying processes that influence customer behavior, improper demand prediction and unconstraining, and optimisation that runs slow and without explicit cancellations.

In Chapter 2, the traditional definition of RM is introduced: "selling the right product (or seat) to the right customer at the right time to the right price" [1]. We claim that this definition has limitations (inability to target the "right customer"), lacks elements (capacity, distribution channel, customer behavior, cancellations, total revenue), and is ambiguous (right seat, right time, right price, and does not mention *how* this is solved). Instead, we propose a new definition of RM, which is: *dynamically assigning capacity to products of perishable nature with a fixed total capacity*. This is underpinned by the science and practice of segmenting your customer base, construct products and price them accordingly, forecasting true demand and cancellations, determining a time-dependent optimal policy that maximizes total net revenue under some given capacity, by modelling internal (customer-facing) and external (market-facing)underlying processes, keeping in mind

practical technical limitations. This is followed by a review of the theory of demand prediction and pricing.

The practice of revenue management is studied next. In Chapter 3, exploitative customer behavior and the "smart customer" is discussed. It is shown that there is no financial incentive for changing or cancelling tickets until one hour before departure, the "smart customer" will therefore wait until the very last moment. From a exploit perspective, seasonality, stopover, point of sale, one-way and minimum stay abuse and nesting bookings is explained and potential savings to the customer are illustrated using real data. In Chapter 4 technical limitations of the RM system in practice are discussed. These are grouped into demand forecasting, pricing, optimisation and distribution challenges that are not covered in the literature.

Next, forecasting demand and cancellations is studied. Once a booking class is closed for sale, an airline does not record any bookings for that class and thus needs to uncensor this value to obtain a true demand estimate. This process is called *unconstraining* and is discussed in Chapter 5. It is shown that the proposed method works well when data exhibits nonlinear inhomogeneous Poisson rates, interarrival times that may not be exponential and discontinuities in demand. It is also shown that this novel method of unconstraining works well in situations where a large number of data points are constrained, which is a drawback of traditional methods. In Chapter 6, a new method is introduced for itinerary-level booking prediction. This is achieved by studying both static elements, such as quality of in-flight entertainment (IFE), as well as time-dependent, underlying processes, such as fare movement. After setting hyperparameters, extreme gradient boosting was used to identify what drives decision-making in ten different ODs (origin-destination pairs). Using empirical data, it was shown that ODs can be grouped into price-sensitive ODs (driven by price), departure time sensitive ODs (driven by departure time), flying comfort ODs (driven by quality of in-flight entertainment).

Predicting cancellations is discussed in Chapter 7. It is shown that cancellations are driven by three different, time-dependent processes: ticketing deadline, fare movement and departure time. Moreover, it is made evident that the time of cancellation depends on the time of booking. The objective is to forecast time of cancellation, depending on the time the product was booked. A three-step framework is introduced that consists of (i) identifying whether a booking is cancels, (ii) categorizing the type of cancellation, and (iii) using probabilities found through Bayesian Inference to predict the time of cancellation.

In Chapter 8, the concept of downsell and postponement of decision is discussed. Downsell occurs when a customer purchases a lower fare than she originally intended. Customers that postpone their decision may find a cheaper class available than was originally available. Two new dynamic programming formulations are proposed: in one, it is assumed customers will always purchase the lowest available fare. In the next formulation, a constraint is added that once a booking class is closed, it may never be opened again. This ensures non-decreasing class availability, and therefore postponing a decision from a customer's perspective is worthless (fares will never drop). Increases in revenues are shown, and it is argued that this will stabilize customer behavior, stabilize bookings, and in turn ease forecasting, too.

Chapter 9 provides optimisation with cancellations through a heuristic. The optimal policy does not compare the exact fare, but rather an adjusted fare is evaluated against the opportunity cost. This adjusted fare is calculated based on the different probabilities of cancellation time, which were competed in Chapter 7. Revenue gains are reported, and it is shown this heuristic is robust against over- and underforecasting, as well as predicting cancellation probabilities too early or too late.

1 INTRODUCTION

This thesis covers the real practice and theory and revenue management (RM). We place an emphasis on real, since the vast majority of publications in literature do not consider practical implications, despite claiming to do so. Current literature either oversimplify the RM problem, provides solutions that take long to compute, solutions that cannot be implemented in practice or provide solutions that cannot be explained to an analyst. In my career, I have worked for two different airlines, and worked with five hotel chains, and have seen how RM works in practice. In this introduction, it is our objective to expose some of these limitations. In later chapters, we provide a method to tackle some of the risks we raise.

To understand these limitations, consider a RM department in practice in airlines. In airlines, there are pricing, demand and inventory teams. The pricing teams are responsible to set fare levels. The demand teams are responsible to forecast demand given these fare levels. Finally, inventory teams are responsible managing the flight's availability and overbooking levels.

1.1 PROBLEMS WITH FORECASTING

In practice, RM departments frequently have competing key performance indicators (KPIs). This is often a result of how a department is structured. The demand team is responsible for ensuring an accurate demand forecast, one of the key inputs for the RM system. Do note, however, that we used the word "accurate", not "optimal". For brevity, we avoid a discussion on what an "optimal" forecast means, but literature by Weatherford and Belobaba [2], for example, shows that *overforecasting* results in higher revenues. In practice, accuracy is typically measured as the difference between (historical) forecast and the actual, produced, demand. This itself exposes a conflict of interest within the demand team: the analyst is concerned with getting their forecast as close to actual production while the airline would benefit from a positively-biased forecast.

Next, consider the implications of an origin and destination demand forecast. Consider an airline that serves 100 airports, and operates two daily flights between each pair, say a morning and afternoon flight. The airline only has one particular aircraft with 200 seats. This means that in theory, there can be $100 * 99 = 9900$ origin and destination pairs that this airline offers. Every origin and destination pair has at least two points-of-sale. [1] To produce an suitable forecast, the day is split between before and after noon. Since every pair is served by two flights, there is a total of $2 * 2 = 4$ different possibilities to travel between origin and destination pairs. This results in the need for a forecast for $9900 * 2 * 26 * 4 > 2 * 10^6$ different products, while only offering 400 flights per day, or $400 * 200 = 80000$ seats. If an airline has ten different data collection points (DCPs, a range of time over which demand is aggregated to reduce dimensionality), this represents an average forecast of 0.004. The average size of forecast also implies likely very high estimation errors. In many papers, a forecast is assumed to be given. The example above illustrates that an accurate, robust forecast is very challenging. In research that studies demand forecasting, demand is often estimated based on origin and destination, sometimes including point-of-sale, but in absence of class, time of day and flight sequences. Note that the example disregards the concept of booking classes. Booking classes, represented by the letters of the alphabet, are the products an airline offers. The end objective of RM is to determine how much capacity to allocate to each booking class. For this reason, demand forecasting is extremely important.

There have been developments in demand forecasting and unconstraining. Most of these methods are based on *expectation maximization* or *projection detruncation* algorithms. These methods are discussed, for example, in Brummer [3] or Guo et al [4]. While both methods work well, the methods suffer lack of performance when one has only very few data points. Consider again the example above, with an average forecast of 0.004. This indicates a very large number of observations with a demand of zero, far outweighing the number of positive observations (which are required to unconstrain properly).

There may be some business rules that restrict certain booking classes for sale based on some condition. For example, a popular heuristic is to close a booking class when a certain load factor threshold is reached. If this threshold is usually reached early in the booking curve, there are never observations for this class in times further into the booking curve. This makes traditional unconstraining impossible, as there are literally no observations to use. This calls the need for an alternative unconstraining technique.

Another approach to forecasting and unconstraining is *Q-forecasting*, introduced by Hopperstad et al [5]. This is a multi-class method that first forecasts the lowest class, then distributes this demand across higher fare classes. This method depends on a willingness-to-pay curve, or similarly, an estimate of the ratio for which passengers are

1 To see why a origin and destination pair can have more than two points-of-sale, consider an airport such as Basel. This point of origin is offered by Germany, Switzerland and France because of its geographical location.

willing to buy a higher fare. This estimate, the fare ratio for which 50% of passengers will purchase a higher fare, the so called FRAT5, suffers from one main drawback: it is highly subjective. In practice, true sell-up is very difficult to measure. And, as it turns out, the revenue is heavily impacted by the choice of this parameter. To illustrate why this is difficult, consider the following example. Suppose an airline offers three classes, 1 through 3, decreasing in fare. Assume that all classes are available for sale. If a customer purchases class 2, it is clear that this customer is specifically looking for class 2, because class 3 would have been cheaper. However, now suppose that class 3 is closed. Once more, a booking for class 2 is recorded. But is this a "true" class 2 customer, or class 1 customer that wanted to pay less, or a class 3 customer that was forced to pay more? This calls for a new method for unconstraining and demand estimation.

Most demand forecasting methods are *internal* facing. That is, they tend to use the airline's own data, and use these to make predictions for the future. The "why" is often not considered. Everything happens for a reason, but the underlying processes that drive customer choice are not considered.

Most of the research in forecasting is focused on demand forecasting, while cancellations have received much less attention. This is concerning, since cancellations pose a big risk for the airline, taking up valuable units of capacity. When cancellations are forecasted, they are often based on some level of aggregation. In practice, the same applies: demand forecasts are looked at in great detail, while cancellation forecasts are often overlooked or ignored.

The current theory of *forecasting* cannot be applied in practice since:

1. The large number of ODs mean very limited data points to base forecasts on, despite levels of aggregation, causing forecasts that are not robust.
2. Cancellations are typically done at an aggregation-level, causing both large errors and variability.
3. Traditional unconstraining are unreliable for smaller ODs at best and impossible to compute at worst, resulting in inaccurate unconstrained demand forecasts.

1.2 CHALLENGES WITH PRICING

Pricing teams are tasked to determine how to position products: a product is a combination of fare conditions and a corresponding price.

In practice, little research is done in either positioning or setting fares. Rather, analysts often only look at what the competition is doing, look up their fare, and arbitrarily add or deduct an amount, or simply match it. This illustrates how irrational pricing can be. Pricing teams determine fare levels, but not every product may be available for sale. After all, this is the very output of the RM optimization problem: determining how much to sell of every product. Yet, pricing and sales teams are often measured by how many products they are able to sell.

While demand and pricing teams cover origin and destination pairs, inventory teams are responsible on a flight level. The inventory teams not only determine to what extent a flight will be overbooked, but also what booking classes are available for sale. This introduces another competing KPI: an inventory analyst may decide to close a booking class for which a pricing analyst has filed a fare that she thinks is able to sell well.

Another problem in practice is the severe distrust in RM systems. This often happens when a change in systems is made. Change management has extensive literature. But even in absence of a change in systems, analysts often think they can outperform the RM system, despite all its complexities. Analysts often cling onto a single day or flight where the system gave a - what appears to them - suboptimal recommendation. As a result, they override the system and take credit for this change. They do, however, forget the other 364 days in the year where the system did give a good recommendation. However, there is another danger to this: changing the system also changes the history. Consider, for example, an inventory analyst closing a given class. Then, since this class is closed, there is no more demand recorded for this class. Often, there is no feedback loop between the RM system and inventory system. This, in turn, means that the RM system does not uncensor demand for this class. And this, in turn, means lower demand estimates for future departure dates. Lower demand means better availability. Better availability means more seats available for lower booking classes. And this, finally, means lower revenues.

As mentioned above, setting prices is usually done based on intuition. There have been several papers in the literature that discusses how to set optimal pricing levels, such as Zhang and Weatherford [6]. Studying pricing under customer choice was recently studied by Gallego and Wang [7]. Marketing from a customer perspective, and the fairness of RM is discussed in McMahon-Beattie et al. [8] and Yeoman et al [9]. However, there has not been any research into how current pricing can be exploited by the customer. We define an *exploit* as a way for a customer to pay less for their journey than the airline intends to, through a loophole. These exploits range from incorrect availability - giving the customer the ability to buy a lower fare class than they were willing to pay for - to other methods that exploit fare rules. While we agree that defining price structures is important, if there is substantial abuse one may question the purpose of these pricing structures.

This leads us to the following conclusion:

The current theory of *pricing* does not work in practice because:

1. Pricing is often done by looking at competition, while research is most often focused on price elasticities, in optimisation there is no notion of competitors.
2. Distrust in the RM system, while the theory assumes optimisation policies are always followed.
3. Competing KPIs between departments, while the literature tends to focus on maximizing revenue.

1.3 COMPLICATIONS IN OPTIMISATION

Consider optimization techniques. It is undeniable that the exact formulation of the network RM problem suffers from the curse of dimensionality. There has been research that study different approximate dynamic programming techniques and heuristics to reduce the dimensions, and therefore, finding an optimal solution possible. Consider for example the method proposed in a widely-cited paper by Adelman [10]. While Adelman found a way to reduce the number of dimensions, this approach cannot be implemented in practice. The biggest problem in this paper used a total number of time units of $T = 1000$ and just 20 destinations, which took over 373 seconds to run. In practice, we often need values of T in the 100.000s, and a much larger number of destinations. Another approach, published by Zhang et al. [11] faces the same issue. In their largest example, runtimes with values of $T = 800$ are reported that took over 2532 seconds to run. However, this problem only consisted of 336 products. As we discussed above, a mid-size airline may offer over $2M$ products. While one can combine, or aggregate some of these products, it is impossible to get the number even close to the 336 products used in their paper. These aforementioned problems consider the true RM approach, that is, assuming fixed capacity and adjusting availability to obtain maximum revenues. Another approach, through dynamic pricing, published by Ke at al [12], suffers the same fate. Even with a small network consisting of 18 flights, 99 products and $T = 200$ time units, run time already exceeds two minutes.

Most optimization methods disregard cancellations explicitly in their input. Instead, they do so by using "net" demand in the optimization process, that is, the sum of all bookings minus the sum of all cancellations. While this approach works well if there is enough time to resell a seat, it becomes a problem closer to departure. Cancellations very close to departure may cause challenges to resell this seat before the flight departs. Therefore, cancellations should be part of the optimization method. Cancellations are currently Boolean: they either happen, or they don't. It is clear, intuitively, that if it is known when a booking cancels (if it does) this enables the system to make a better recommendation by either overbooking more efficiently, or rejecting this request altogether.

The current theory of *optimisation* does not function in practice as:

1. The number of fare/product combinations is severely underestimated, meaning inputs are often not reliable in practice.
2. Proposed techniques have runtimes that are too long, making it impossible for airlines to reoptimize as frequently as they would like.
3. The time dimension of cancellations is not explicitly modelled in optimisation methods, exposing the airline to unnecessary risk.

1.4 OBSTACLES IN CULTURE

Finally, we would like to highlight the concept of culture in practice. This should noe be underestimated. In the previous sections, we discussed competing KPIs within the same department. We have also seen that an analyst often thinks he or she is able to outperform the system. There is, today, still a significant distrust in the system and its output. In some cases, these concerns are valid: subpar inputs result in subpar outputs. However, even when forecasts are how they should be, problems often persist. This could be explained by culture, limited training, or lack of understanding of how the system functions.

In practice, is it therefore often much more desirable to have an algorithm that performs well and is easy to explain rather than a black box method that performs extremely well. A true measure of how well a method works should not only measure model performance or accuracy, but also trust. A proxy to trust can be found in the number of times the algorithm is overruled.

To build trust in a RM system, solutions should be easily interpreted. This is why one of the first optimization methods, the expected marginal seat revenue (EMSR) heuristic, by Belobaba [13], is still widely used in practice. The solution method is fast, and an analyst is able to understand the output. Methods should not only focus on model performance, both in terms of speed and revenue, but also whether the inner working of the model can be explained to an analyst, that may lack knowledge of mathematics.

So far, we have looked at the airline side of things. However, we should not forget the culture of customers. Some features, a result of the legacy of RM, offer the customer opportunities that expose the airline to risk. Moreover, with the rise of the internet and the emerging information age, customers have more access to data and information than ever before. This has changed the mindset of the customer. It is typically assumed that customers are honest, and do not engage in exploitive behavior. However, airlines have seen a steady rise in abuse, as evidenced by lawsuits (for example [14], [15] and [16]). It is easy to disregard and forget about this side of RM, and be internally (systems) focused, but this is an area that poses risk to the airline and has not received any exposure from the community.

Summarizing, the theory of RM is lacking in practice because:

The current theory of RM does not align with *culture* since:

1. The quality of information is not regarded, perfect information is assumed.
2. Culture in the RM department is overlooked and observation bias by analysts overriding solutions that are not well understood.
3. The customer is seen as "honest", not using exploits.

1.5 CONCLUSION

With this introduction, we give the reader an insight into the complexities of RM in practice. We have seen that difficulties in demand forecasting, pricing and optimisation may all hinder revenue performance. However, one must not overlook the concept of culture, either. Within the scope of this thesis, it is impossible to tackle all of the problems highlighted in previous sections. We have summarized the problems identified in the previous sections and categorized into the three main problems, which are shown in the box below.

The current theory of *RM* does not align with practice because of the following problems:

1. Maximizing revenue is often seen from a systems perspective, focussed on accurate inputs, but puts the customer's behavior on the background.
2. Demand is unconstrained through methods that are not reliable in the OD-world, forecasts tend to use internal data and don't consider the "why", cancellations are not a priority.
3. Optimisation and forecasting methods do not accurately consider cancellations and run too slow to ensure frequent reoptimisation.

The RM problem from a customer's perspective is investigated in Chapter 3. Here, an overview of the legacy of RM systems is given. After this illustration, several exploits are discussed that enable the customer to circumvent fare fences. Next, changing, cancelling, and rebooking is discussed. While these are not exploits, they do pose a risk to the airline. This chapter looks into Problem 1 by putting the customer at the heart of things.

The aforementioned information age has made it extremely easy for customers to compare prices. It may be argued that more and more customers are becoming price sensitive. In fact, most customers will make claims like "the price went down" while in reality, availability has changed and they book a different product. In Chapter 8, a new dynamic programming formulation is given that explicitly models customers that book the cheapest fare an airline offers. This is then extended by considering customers that may postpone their decision and benefit from better availability later on. This formulation considers Problem 1, and by virtue of fast run times, it enables the airline to reoptimize frequently and thus tackles Problem 3 as well.

We have seen that forecasting in practice is much more complicated than it is typically made out to be. In practice, the large quantity of products mean that there are often very few non-zero observations. In fact, there are products that are never sold at certain periods in the booking curve. This means that traditional unconstraining methods are unreliable at best and impossible to compute demand estimates at best. In Chapter 5 a new method for unconstraining is introduced. This tackles Problem 2 above.

We have stated that most of the current work on forecasting uses only *internal* data. We claim that by studying underlying processes that may drive customer choice, we will obtain a higher forecast accuracy, while at the same time understanding the "why". In Chapter 6, we introduce a machine learning solution that considers five underlying processes that may affect customer behavior. We develop a method that is both easily scalable, runs fast, and one that works better with a low number of observations compared to traditional methods. Problems 1, 2 and 3 are tackled.

Cancellations are studied in Chapter 7. A framework is provided for estimating booking-time dependent cancellation probabilities. This is done by considering underlying processes, similar to the ones used to describe booking behavior. The framework focuses on the *"why"*. Problems 1 and 2 are fought.

These cancellation estimates are then introduced in a new dynamic programming formulation in Chapter 9. Demand is forecast using the method described in Chapter 6, unconstraining is done using the algorithm proposed in Chapter 5 and cancellation probabilities are given by the framework of Chapter 7. Therefore, this chapter tackles all the Problems 1, 2 and 3 outlined earlier.

This thesis does not only provide possible solution methods for unconstraining, demand forecasting and optimization, and shows why the model recommends a certain action, but it also provides a unique look into the current practice of RM. This makes this thesis, the **real** theory and practice of revenue management.

1.6 PUBLICATIONS

The results presented in this thesis are based on the following papers:

Daniel Hopman, Ger Koole, and Rob van der Mei. Exploitative customer behavior in airlines: an overview of strategies for the smart customer. Working paper. Manuscript in preparation. This paper is the basis for Chapter 3.

Daniel Hopman, Ger Koole, and Rob van der Mei. Practical limitations of revenue management: theory versus practice. Working paper. Manuscript in preparation. This paper is the basis for Chapter 4.

Ilan Price, Jaroslav Fowkes, and Daniel Hopman. Gaussian processes for unconstraining demand. *European Journal of Operational Research*, 275(2):621–634, 2019. This paper is the basis for Chapter 5.

Daniel Hopman, Ger Koole, and Rob van der Mei. A machine learning approach to itinerary-level booking prediction in ten competitive airline markets. Working paper. Manuscript in preparation. This paper is the basis for Chapter 6.

Daniel Hopman, Ger Koole, and Rob van der Mei. A framework for modelling time-dependent cancellations in airline revenue management. Working paper. Manuscript in preparation. This paper is the basis for Chapter 7.

Daniel Hopman, Ger Koole, and Rob van der Mei. Single-leg revenue management with downsell and delayed decision making. *Journal of Revenue and Pricing Management*, 16(6):594–606, 2017. This paper is the basis for Chapter 8.

Daniel Hopman, Ger Koole, and Rob van der Mei. Leg-level dynamic programming with booking-time dependent cancellation probabilities. Working paper. Manuscript in preparation. This paper is the basis for Chapter 9.

Part I

THE THEORY OF REVENUE MANAGEMENT

2 LITERATURE REVIEW

This chapter provides an overview of the current theory of pricing and revenue management (RM). RM takes demand forecasts, pricing and capacity as inputs. It is typically thought that capacity is given. Therefore, we focus on forecasting, pricing and optimisation. We start by giving the traditional definition of RM in Section 2.1, and show why this definition does not align with the real theory and practice of RM. This is followed by Section 2.2, in which we review demand prediction. Demand prediction may be achieved through surveys (Section 2.2.1), statistical methods (Section 2.2.2) and machine learning (Section 2.2.3). The process of uncensoring booking observations, the concept of unconstraining, is reviewed in Section 2.3. Cancellations are discussed in Section 2.4, which are studied on an aggregated (Section 2.4.1) and disaggregated level (Section 2.4.2). The field of pricing is reviewed in Section 2.5, which consists of customer segmentation (Section 2.5.1), the process of setting the actual fares (Section 2.5.2), and ensuring this policy perceived as fair by the customer (Section 2.5.3). Optimisation is discussed in Section 2.6. This is separated between leg-level (Section 2.6.1) and network-level (Section 2.6.2) optimisation.

2.1 REVENUE MANAGEMENT

What is RM? A quote that is often referred to says that RM is "selling the right product / seat to the right customer at the right time to the right price" [1]. It can be argued whether this is an accurate definition. Below, it is described why this definition is both wrong and incomplete. There also is a difference between the concept of RM for an airline and that for a customer.

First, ignoring any "premium" seats (such as emergency exit rows, that are often charged at a premium), an airline does not really care whether a passenger selects seat 23A, 68A or 80J. Yet, with the amount of different itineraries airlines offer, customers are all likely to pay a different fare, despite being seated next to each other. Therefore, **selling the right seat** is to be omitted from the definition of RM. On the other hand, the passenger, may have a preference to sit on the left, or right hand side of the airplane. From a customer's perspective, this should be included.

Second, consider **the right customer**. This definition implies that an airline is able to segment their product offering in such a way that it is able to target specific customers. However, as will become evident in future chapters, there are two challenges with this: firstly, customer behavior is often erratic and therefore impossible to segment properly. Secondly, customer segmentation is often done manually, without proper (statistical) analysis. Thirdly, and most importantly, as a result of legacy systems, most airlines are bound by 26 products (the number of letters in the alphabet) across their classes of service: one may argue that this is not sufficient to properly segment a customer base.

Third, it is not immediately clear what the concept of **the right time** is. One thing that is for certain is the fact that units of capacity is perishable: once a flight takes off, any empty seat is lost revenue (this, by the way, is the reason why staff standby travel is such a win-win for both the airline and staff). Therefore, requests arriving after a flight is closed for sale are worthless. However, it is not immediately clear what the right time for similar requests before this time is. For example, is a customer purchasing a fare worth $100 40 days before departure any less valuable than 5 days before departure? Interestingly, from an optimisation perspective, the opportunity cost decreases as time draws closer to time of departure, yet it is typically thought that a customer's willingness to pay increases as time closes in. Therefore, while the optimal policy may say that it is the "right time" to sell this cheap product, while practice says it is the "right time" to sell an expensive product.

Fourth, **the right price** is both ambiguous and impossible. Firstly, it is ambiguous because the right price is different for everyone. Even for the same customer, a price today may be perceived as fair, while that same price tomorrow may be perceived as being unfair. This is typically driven by competition: if all competitors charge $200 and one airline charges $150, this is considered fair (and probably even thought of as being a great price). However, if competition charges $100 while that same airline charges $150, this is now considered as being unfair. Secondly, even if one has perfect information and is able to predict willingness to pay with perfect precision, the discretization of price points, a result of only being able to use 26 different levels, means that it is impossible to charge exactly what one would like.

Fifth, this definition lacks the notion of **capacity**. Airlines typically have different teams or departments that decide what type of aircraft (and therefore, capacity) to deploy on what routes. It is often assumed that this capacity is fixed and cannot be changed. Clearly, the RM problem is closely tied with capacity allocation.

Sixth, it lacks the acknowledgement of **different distribution channels**. In Chapter 4, we illustrate that fares sold through an airline's own channel have a different value than fares sold through a travel agent, which is paid a commission. The fare value received after commission and fees is often referred to as the "net fare" and it is important that this is maximized, not the (gross) fare.

Seventh, the definition lacks **customer behavior**. In Chapter 6, we show that customer behavior is different across different itineraries. Therefore, it may be important to define products differently for different itineraries. It is also important to note that legacy systems make exploitative behavior possible, which is shown in Chapter 3 and this isn't captured in the definition of RM above.

Eighth, it does not cover **cancellations**. The definition only covers the art of "selling" a seat, and does not consider what happens when cancellations occur. Making this distinction is even more important because, as we show in Chapter 7, the reasons why customers cancel differs from the reasons why customers booked that itinerary in the first place.

Ninth, it does not consider **total revenue**. RM, along with the planning department, often focuses on creating the best possible schedule. This is achieved by having one or more "arriving / departing" banks that connect traffic flows. For example, KLM's flights from Amsterdam to Asia depart between 8 and 9M. They then ensure that flights from Europe into Amsterdam arrive around 6PM. The time between 6 and 9PM is then known as an "arriving / departing" bank. While this is great for the customer, this creates a strain on the airline during this time period: flight delays into Amsterdam may cause customers to miss their connecting flight. It is then the airline's duty to rebook a customer at the earliest possible connection, even if this means rebooking onto other airlines. This, of course, reduces the *total revenue* the airline receives. While a tight schedule may be beneficial for customers, it may not be beneficial to maximize total revenue.

Tenth, it does not show **how this is done**. This is important because it should be *possible* to solve the definition. It should not only be possible from a mathematical perspective, but it should also be possible to be implemented in practice. In Chapter 4 we illustrate limitations in practice.

Therefore, it is concluded that the definition of RM of Cross [1], has both limitations (right customer), lacks elements (capacity, distribution channel, customer behavior, cancellations, total revenue), and is ambiguous (right seat, right time, right price, does not show how it should be solved).

Instead, the following definition of RM is proposed, that addresses these shortcomings:

Definition of RM
Dynamically assigning capacity to products of perishable nature with a fixed total capacity.

This is underpinned by the science and practice of segmenting your customer base, construct products and price them accordingly, forecasting true demand and cancellations, determining a time-dependent optimal policy that maximizes total net revenue under some given capacity, by modelling internal (customer-facing) and external (market-facing) underlying processes, keeping in mind practical technical limitations.

2.2 DEMAND PREDICTION

Beckmann [24] provides the first framework of RM. In this work, published in 1958, he claims that reservations are a form of rationing. This is followed by his premise: "what is the rationale of rationing in a market economy premised on the principle of distribution through price competition rather than direct rationing?" [24]. When it comes to demand, he then asks: "should it stop reservations at full capacity or should stop later, anticipating a certain number of later cancellations and 'no-shows'"? Using data provided by an unnamed airline, he shows that demand can be modelled using the Gamma distribution. This is followed by finding an optimal policy, but this is outside the scope of this section: the key take-away here is the ability to model demand using Gamma distribution. Beckmann, followed by Littlewood [25], are widely credited as being the first to describe the process of optimisation.

However, the key assumption that both used, and many authors use up to this day, was the assumption of statistical independence of demand for products. A product is the set of flight characteristics, associated fare, and fare conditions offered to the customer. Flight characteristics include departure date and flight number. The associated fare is the fare that is offered to the customer. Fare conditions include, among others, cancellation and change policies (free, at a charge, or not allowed). In effect, if a customer is forecasted to purchase a specific product *p*, that customer will always purchase *p*, and never any other product available either by the same airline, or in the market. Throughout this Thesis, this is what we refer to as being "independent demand". Demand is assumed to be independent both for obtaining demand estimates as well as heuristics that maximize revenue.

Customer-Choice Revenue Management (CCRM), on the other hand, is different. Here, it is assumed a customer is given a set of options and chooses accordingly. These alternatives may be offered by the airline itself - for example, a more suitable departure time; or by its competitors - for example, a lower fare. For further background, we refer the reader to Poelt [26], who provides an excellent description of the underlying process and analysis of this shift. Poelt discusses how airlines have traditionally segmented customers between business and leisure travellers: business travellers are less price-sensitive, less flexible, sensitive to schedule times, book later, and travel less. Leisure travellers are considered the opposite. Poelt then shows the different inputs airline may use for their demand forecasting process and illustrates the dangers of independent demand forecasting, that is, without dependency between different choices offered to the customer.

In what follows, we separate between findings from surveys and underlying processes, modelling through statistical methods, and modelling through machine learning techniques.

2.2.1 *Surveys and background*

In this section, we provide an overview of surveys and the background of the decision making process for bookings. These surveys are conducted in different countries, and aim to illustrate the rationale behind making a booking. The importance of loyalty programmes is also discussed. Other industries, including hospitality are also discussed.

Woodside and MacDonald [27] provide a framework for the decision process making as a whole in the tourism industry. From the initial research, to travelling to the destination, to how time is spent at the destination, to leaving: Woodside and MacDonald conducted interviews that map this decision making process for visitors to an island in Canada. They discuss demographic, psychological, personal values; influence of family, friends and groups; marketing. This is followed by the search for information, the evaluation of this information, and the use of a heuristic to make a decision. This decision is driven by destination, destination area, activity, accommodation, attraction, route, dining and shopping choices. One of their findings that the decision of travelling to the destination was made on a separate day than travel plans to that destination were made. This delay in decision making is discussed in Chapter 8.

The concept of loyalty in the aviation and hospitality industry is very important. After all, it is agreed that attracting a new customer is much more expensive than retaining a current customer. Most airlines have loyalty programmes. They claim that these programmes offer substantial benefits to customers: they allow customers to collect miles for every mile flown and spend these for free flights or upgrades in the future. However, this is an excellent tool for airlines to track customer booking behavior. Therefore, the booking behavior of loyal customers is studied. Of course, this calls for a definition of a "loyal customer". Airlines with frequent flyer programmes only capture data of flights they operate themselves: they have limited to no visibility what other airlines their customers use. The definition of the "loyal customer" is outside the scope of this work. The behavior of "loyal" customers is studied by Dolnicar et al. [28] They do not define what a "loyal" customer is, so instead we propose the term "frequent flyer". Through a survey with almost 700 responses, they show that the frequent flyer programme, price, national carrier and the reputation of the airline through word of mouth are the four key drivers to discriminate airlines. Interestingly, they show that customer satisfaction was not a driver to airline loyalty. They also note that this conclusion only applies for business travellers, and claim it is difficult to model this for leisure travellers.

Dowling and Uncles [29] investigate whether loyalty programmes are a driver for success. They claim that customer choice is based on the category purchase decision ("do I fly or take the train?") and the brand choice ("do I fly British Airways or Air France?"). They claim that airline should focus on high engagement with the customer to ensure that both of these choices are made in their favor. According to the authors it is hard to obtain advantages through a loyalty programme and any benefits are quickly overshadowed by competition.

Sandada and Matibiri [30] study the service quality and presence of a frequent flyer programme in the airline industry in South Africa. Through a survey conducted with 148 respondents, they study the effects of customer satisfaction, service quality, frequent flyer programmes and safety on loyalty. They find that customer satisfaction and a frequent flyer programme result in customer loyalty. They also show that the safety of the airline and its reputation is not significant. While they argue that a customer programme is important, they also argue that having satisfied customers alone is not enough.

The differences between full service, legacy and low cost airlines is discussed by Koklic et al [31]. Through a survey conducted in the European Union, yielding 382 responses, they aim to study what characteristics influence customer satisfaction, intention to repurchase and intention to recommend. They show that staff positively influences satisfaction and satisfaction positively influences the likelihood of a customer returning. The quality of staff is of greater importance in legacy airlines, while the effect of satisfaction on repurchase is greater for low-cost airlines.

Tsikriktsis and Heineke [32] discusses the impact of customer dissatisfaction in the domestic US airline market. This provides light on what drives booking behavior. They introduce four major factors that contribute to variation in service delivery: lack of well-defined processes, high employee turnover, heterogeneity of customers and customization. The two main hypothesis are better average process performance, as well as lower variation in these processes, lead to higher customer satisfaction. A distinction is made between high and low performers. They find that both of these hypotheses are valid, but only for high performers. This seems to indicate that service delivery should be important to an airline, not just wanting to be the cheapest. They conclude that consistency is just as important as average performance.

Park et al. [33] go one step further, and move from the high-level work of Tsikriktsis and Heineke [32] to study the individual dimensions of airline service quality affects customer choice. This was done through a survey conducted of passengers in the international Australia market. They attempt to model what characteristics customers associate with a positive brand image. Analysis of surveys show that that in-flight service, convenience and accessibility were each found to have a positive effect on airline image. Furthermore, they show that there the main relationship that drives behavior is the relationship between perceived price and perceived value. They then find that these characteristics were directly related to future decisions of these customers.

Carlsson and Lofgren [34] study the domestic market of Sweden over a time period of ten years and discuss the cost of a customer churning. While the objective of this paper is to study what the cost of a switching customer is, they present findings that show why customers book with competitors. A first finding is the importance of competition. Next, for this domestic market, the number of daily flights for the airline, as well as their competitors, is statistically significant in the decision-making process. They study two airports: a smaller airport closer to the city centre, and one large airport away from the city centre. The convenience of being close to the city centre also proves to be significant. They also investigate the effects of a frequent flyer programme: interestingly, only one particular programme is statistically significant, which seems to indicate that not every frequent flyer programme is perceived of the same value when deciding what airline to choose.

The domestic market of India was discussed by Khan et al [35]. They make the distinction between public and private full service airlines, and low cost airlines. For their analysis, they focus on two categories: in-flight services and support services. Through a survey, they identify customer expectations, perceptions and the gap. For in-flight services, it was found that customers flying full service airlines were happier with cabin ambiance and cleanliness. For support services, it was found that the smaller networks low cost airlines offer have the biggest gap. In terms of reliability, they find that full service airlines are poor at handling irregular operations (IRROPs), while low cost airlines are perceived to perform better. Expectedly, the gap between expectations and perception in customer interaction and personalization onboard was found to be greater for low cost airlines compared to full service airlines. Related to this, the gap in responsiveness to customer requests, was much greater for low cost airlines compared to full service airlines.

Atalik and Arslan [36] study the the Turkish market. Through a survey with 397 respondents, they study what characteristics matter. They find that 80% of customers are satisfied with the value-for-money proposition they were offered at time of purchase. Of those 20% that were not, it was identified that they felt airlines did not properly communicate what their product offering was. It was found that customers place importance on airlines providing a communication channel for suggestions to improve this. When deciding what airline to choose, it was found that flight safety was rated highest, followed by on-time performance, staff and image of the airline. Only after these variables, the price of the ticket was listed in terms of importance. Compared to other studies, this seems to indicate that different markets put different levels of emphasis on fare value. Catering, interestingly, was found to be least important.

Cho [37] studies the domestic United States (US) market. Through surveys, he investigates the choice of airline and airport. In the US, many cities have more than one airport. For example, the city of New York has three major airports, the city of Chicago has two airports and the greater San Francisco area has three airports. This shows that not only choice of airline, but also choice of airport have an influence on decision-making. Cho's first finding is that operational quality positively influences a customer's decision to book a specific airline. Secondly, he shows that greater customer exposure to service offered improves the same probability. Therefore, it is suggested that customers are sent individual offers. In terms of airport choice, he finds that airports that have a low cost carrier presence are preferred, and the proximity of the airport to the city centre is important. He shows that different customers react in different ways: customers that prioritize on-time performance in their decision-making process are affected by operational quality and airport choice, while customers that prioritize price do not.

Baker [38] studies the domestic US market, and investigates the difference between legacy and low cost carriers through data of fourteen US airlines. A perceived service quality model is introduced, consisting of word of mouth, personal needs, past experience and five dimensions of service quality: reliability, responsiveness, assurance, empathy an tangibles. An expected service and perceived service level is than calculated and compared. Interestingly, he shows that the perceived service quality of low-cost airlines is higher than those of legacy airlines, and discusses how the effects of operating costs, market share, and infrastructure influence this metric of perceived level quality.

Keiningham et al. [39] discuss the effect of service failures on the future behavior of customers in the domestic US market. They separate "failures" into "major accidents", those that cause physical harm (resulting in injury or death), and "minor incidents", those that do not cause harm (for example, delayed luggage). In other industries, they claim that the generally accepted view among managers is that the greater the severity of service failure, the greater resulting impact on customer satisfaction and customer choice. They conclude that this does not apply in the aviation industry. They find that major accidents have *less* impact on future market share, than minor incidents. Furthermore, major accidents did not show any impact on customer satisfaction, while minor incidents greatly negatively influence future customer satisfaction. They also find a (insignificant) negative relationship between customer satisfaction and market share.

Customer choice in the Nigerian market is studied by Adeola and Adebiyi [40]. Through a survey conducted with 200 respondents, with different reasons for travel, frequency of travel, occupation and educational characteristics, they study the influence of service quality in the decision-making process. They find that the fare is the most important factor, followed by social status and airport quality. With limited options available in the Nigerian market, they argue that customers would prefer alternative methods of transportation, but the lack of infrastructure forces customers to fly, and are forced to fly airlines with suboptimal service quality.

Khan and Khan [41] discuss the market in Pakistan. Through surveys, they aim to model customer choice based on reliability, responsiveness, assurance, empathy and tangibility. Reliability is studied through on-time departure and arrival, problem solving and (mis)handling of luggage. Responsiveness is investigated through fast ticketing, airport assistance, staff friendliness and luggage arriving quickly. Assurance is measured through brand image, safety record, competent employees and service level. Empathy is studied through understanding special needs, anticipating future problems and individual customer choice. Tangibility is measured through aircraft age, cabin crew, queues and reservation offices. It is shown that all of these, except the location of reservation offices positively influences a customer's decision to book an airline. Assurance and empathy factors have the largest weights.

Hussain et al. [42] study the United Arab Emirates market and, in particular, how service quality affects customer booking behavior. They look into different categories: corporate image, perceived value, customer expectations and service quality. Through a survey with 253 responses, with a variety of nationalities and backgrounds, they study how these categories affect four outcomes: expectations, perceived quality, perceived value, satisfaction and brand loyalty. Corporate image attributes have statistically significant influences on all of these outcomes. Perceived value is shown to have a direct impact on customer satisfaction. Interestingly, they show customer expectations has a direct impact on perceived quality, but not value.

Law [43] investigates the Thai travel market through surveys at four different airports, yielding a total of 600 respondents. The majority of respondents agreed on looking to purchase the lowest available fare in the market. Law shows that the loyalty programme is not important to the respondents. From a service perspective, customers are seen to find in-flight products more important that in-flight services. Interestingly, there was no consensus about in-flight entertainment. This could be because of the nature of the market (mainly flights with short duration). Departure time is shown to be important, while the ease of booking a ticket is shown not to be important. Comparing all variables, Law finds that price is the main driver, followed by schedule. If price and schedule is comparable, only then a customer looks at amenities and services. The loyalty programme is of least importance.

The Indonesian market is explored by Manivasugen [44]. In this research, he asked 140 students about their travel preferences. The most important factors when deciding between airlines were identified as price, comfort, safety, schedule and airline image. On-time performance and baggage services were reported as being moderately important while food and drinks, aircraft type and cabin services were not important in deciding among alternatives.

With increased security measures in airports, and the rise of high-speed trains, airlines no longer compete with peers, but also with other forms of transportation. This will affect how customers make their decision. The growth of the air transport industry as a whole is discussed by Bieger et al. [45], and is followed by an analysis of the perceived value of the customer value. The importance of different characteristics of air travel is discussed after conducting surveys, and is segmented between economy and business class. It is concluded that ticket price is most important for both economy and business class passengers. Interestingly, the number of stops is more important for economy passengers than for business passengers. On the contrary, the presence of a frequent flyer programme is not important to economy class passengers, but is important to business class passengers. Other factors, such as travel comfort, safety, travel comfort and in-flight services are rated equally.

Before demand can be forecast, customer segmentation is vital. As mentioned in the introduction of this section, Poelt [26] mentions how traditionally airlines segment customers between leisure and business customers. Teichert et al. [46] conducted a survey and received over 5800 responses and find that this traditional segmentation no longer captures true customer preferences. A first finding is that class of travel does not determine reason of travel: they find customers that travel in business class for non-business reasons, and customers in economy class travelling for business reasons. Instead, they define five different segments: efficiency, comfort, price, price/performance and all-round performance. They also make a distinction between segments when it comes to reason for travel. For example, it is claimed that the first segment, "efficiency" are mainly frequent flyers that want on-time performance when travelling for work. However, when these customers fly for leisure, they change segments. This example illustrates why traditional segmentation, both for marketing and optimisation purposes, is no longer sufficient.

The hospitality industry has many similarities with the aviation industry, especially in terms of RM: a customer is given a number of choices, each with a different quality of service. A customer needs to decide what is the best value proposition for them. Verma and Plaschka [47] discuss customer choice modelling and its implications in the hospitality industry. They show the progression customer choice has made in this industry and how this has changed the managerial decisions that are being made. Respondents are given a picture of the hotel, a description of room interior, ambiance, dining, room services, accessibility and price. Personalized, on-demand service and brand image are found to be most important. Verma [48] continues to review this booking behavior. Verma discusses how to develop choice models from revealed preference data, for example, from an online travel agent. Next, he discusses how this can be accomplished through stated preference data, through surveys. Verma offers advice how surveys should be conducted and how revealed and stated preference data can be combined.

Customers have access to different booking channels when purchasing a ticket: for example, one may book with the airline over the phone, through their website, through a brick-and-mortar travel agent, an online travel agent, a corporate booking portal, and more. While the airline ticket prices are consistent across booking channels, the fare received by the airline is different. This is studied by Brunger [49], and he refers to this as the "internet price effect". He finds that customers purchasing identical itineraries online are offered a lower fare than those that book through traditional agencies, and show that this effect represents 3 to 8% in terms of fare value. This reinforces the need to include distribution channel in the definition of RM, which we introduced in Section 2.1.

With the rise of the Internet, websites like TripAdvisor enabled travellers to rate and review hotels, in an unbiased way. It is assumed that ratings on this website have significant impact on customer choice: hotels often do anything they can to improve their rating. One example of this is hotels actively responding to reviews: at least one major, US-based hotel chain only responds to negative reviews in an attempt to engage in service recovery. The airline industry has a similar website, Skytrax. Unlike the ratings that determine a hotel's performance on TripAdvisor, ratings are set by the company Skytrax themselves. These ratings are studied by Perez [50]. Perez shows that the score assigned by Skytrax has a weak relationship to the score assigned by reviewers on the same website. Comparing to the hospitality industry's equivalent, TripAdivsor, this could possibly indicate that this the company assigning these ratings may be biased.

Traditionally, it was thought that customers purchasing tickets in premium (first and business class) cabins have a different thought- and decision-making process than customers in economy class cabins. With many airlines reducing, or completely eliminating first class, the interest in the product offering for the business class cabin has recently spiked interest. This is reviewed by Boetsch et al [51]. Through a survey of over 682 respondents, with different frequencies of travel throughout the year and reason for travel, they find that emotional value is of most importance. In fact, brand image is more important than both sleep quality (flat seats) and price paid. This is true for the twelve different airlines that were studied. Brand image was most important for an Asian-based airline and least important for a European-based airline, but still the most important factor. The presence of a frequent flyer programme was again showed to be least important.

While the airline industry, and, followed by the hospitality industry, have been successful at the adoption of RM, other industries are now using similar concepts. This is studied by Currie [52]. Currie claims that every RM system should incorporate features of both the industry, and the Company running it. Tour operators are first discussed, and it is discussed how they face multiple capacity constraints (hotels *and* airlines), the number of combinations in packages and highly seasonal demand. To incorporate RM techniques for cruises, Currie sees challenges in multiple capacity constraints (cabins and life boats), the need to emphasis cross-selling ancillaries and highly seasonal time-of-booking.

2.2.2 *Modelling through statistical methods*

In this section, we discuss the modelling of demand forecasts through statistical models. The majority of these methods are based on multinomial logit models.

Coldren et al. [53] study the classification of itinerary through a multinomial logit (MNL) model in the US domestic market. They work together with an airline and use real data. The MNL model was based on level of service, number of connections, connection time, distance, brand image, fare, aircraft type and time of day. In terms of level of service, the authors find strong evidence that passengers prefer to avoid connections, which results in the inconvenience of changing planes and a higher probability of delay. Regarding connection time, the MNL model shows that customers strongly prefer the best connection, and any other connection is penalized by an extra 45 minutes (on top of the additional connection time compared to the best connection). In terms of distance, it is found that customers have a strong preference to minimize distance. Brand image is found to be an important factor. As expected, higher fares are less desirable than low fares. It is shown that customers have a strong preference to mainline, larger jets. In terms of time of day, itineraries departing before 7AM or after 7PM are not preferred. Interestingly, these are traditionally thought of being desirable for business travellers.

Vulcano et al. [54] study the feasibility of customer choice modelling and the effects of customer choice RM over traditional RM. The choice attributes include base fare, time of day and day of week. They use a MNL model, and introduce a maximum likelihood estimate to parameterize these attributes. Through simulations, they show the differences on revenues between traditional forecasting and their novel way. Revenue potential of up to 5.3% is reported, which are mainly a result of different optimal policies. However, they argue that these results should be taken cautiously.

Lucchesi et al. [55] evaluate customer preferences for a domestic route in Brazil. They do so by comparing multinomial logit, mixed logit random coefficients, multinomial logit incorporating systematic variations of preferences, mixed logit error component and mixed logit error component incorporating systematic variations of preferences. The heterogeneity of preferences in discrete choice models is captured through the inclusion of systematic variations of preferences. The authors claim, however, that other factors influencing decisions are unobservable or difficult to measure. Therefore, they propose to abandon the traditional MNL models typically used. Instead, they find the mixed logit to perform best. They show that customers loyal to different airlines have different levels of loyalty and cost of switching. They also show that customers that pay for their ticket out of their own pocket have a greater price sensitivity than those that do not. This reinforces an assumption that is generally used in the airline industry: charge high amounts to those that travel for work, since their employer will pay regardless of fare level.

Milioti [56] does not use a MNL approach, but rather a multivariate probit model. A multivariate probit model is an approach that simultaneously estimates the influence of (independent) variables on more than one dependent variables. The dependent variable represents positive or negative responses to each variable when considering itineraries. These variables include purpose of trip, final destination, booking method, cost of ticket, mode of transportation, and trip-specific characteristics such as fare, flight schedule, frequent flyer program and in-flight entertainment. They conducted a survey and received 853 responses. Just like research in other parts of the world, they conclude that fare is most important in the decision-making process, followed by the airline's image. Out of those passengers that are not price sensitive, they find that men and business travellers are least likely to be influenced by the level of airfare. They also note that these variables have differences across different socio-demographics and trip characteristics, such as short or long flights.

The level of service quality is often seen as an important proxy when estimating demand, which is something that can be concluded from Section 2.2.1. Chen et al. [57] argue that the level of service is hard to quantify: words such as "good" or "bad" to describe service quality are ambiguous. For this reason, they use another approach to demand forecasting and use fuzzy logic on in-flight service feedback for a domestic airline in Taiwan. They find that proactive cabin crew is most important when calculating service quality perception.

Ratliff and Gallego [58] use customer-choice modelling for a different application. They introduce a decision support framework for evaluating sales and profitability impacts of fare brands by using a customer choice model. Fare brands are a collection of products with identical fare conditions, with the only differing factor price. For example, suppose an airline has three products with free changes and cancellation, priced at $100, $75, $65. These three products belong to one fare family. Different fares are used to create upsell possibilities. In practice, these fare families are often generated without statistics or science, but based on rule of thumb. Through automatically generated surveys and an associated genetic algorithm, Ratliff et al. show how to find the optimal set of fare product and fare families.

Airlines only observe bookings if tickets are available for sale. Once the airline stops selling tickets, every demand request is rejected. Estimating this true demand is the process of unconstraining (or, uncensoring). Haensel and Koole [59] use a statistical way to estimate unconstrained demand and use real airline data from an airline in The

Netherlands. They show that their expectation maximization method (EM) converges quickly, after 10 - 15 iterations, and provide very good estimates: very small errors are reported on fare class-level, especially heavily-used fare classes.

2.2.3 *Modelling through machine learning*

In this section, we give an overview of the research of modelling demand and making predictions through machine learning.

Mottini and Acuna-Agost [60] expose the drawbacks of using a statistical model, specifically the MNL logit model, which was outlined in Section 2.2.2. They state that the MNL model is the most frequent model used in practice. However, despite its popularity, it possesses multiple weaknesses: first, it only considers a linear relationship between variables. Secondly, it cannot determine the order of attractiveness of different options, but, rather, can only identify the most important option. Thirdly, the model suffers from the Independence of Irrelevant Alternatives (IIA) property. This property states that if choice 1 is preferred to choice 2 out of the choice set $\{1, 2\}$, introducing a third option 3 (thus expanding the choice set to $\{1, 2, 3\}$) cannot make 2 preferable to 1. Instead, they use a machine learning approach typically used in machine translation, called pointer networks. Pointer networks work with two functions: an encoder and decoder. Mottini and Acuna-Agost use Recurrent Neural Networks for the encoder and decoder for hidden states, which means the probability of an itinerary can be calculated given characteristics without making assumptions on statistical independence between variables. The main benefit of a pointer network is the ability to have a variable length of input. In this context, this is great since customer choice sets may have different lengths. Features that were used include distribution channel, airline, Saturday night stay, price, trip duration, departure weekday, number of connections and number of airlines. It was shown that based on this model, and these features, the ML and MNL algorithms were outperformed.

In a similar way, Lhertier et al. [61] discuss itinerary choice modelling. They categorize two kinds of features: features that describe the individual (which they call characteristics) and features that describe alternatives (which they call attributes). They first introduce the MNL, and explain how a customer's utility may be calculated. They then follow by introducing the latent-class MNL model, which takes into account individual heterogeneity by considering different classes in which homogeneity is assumed. This statistical approach has certain drawbacks, which were mentioned above. Therefore, Lhertier et al. introduce a supervised machine learning algorithm called random forests. Random forests is an extension of decision trees. It is well known that decision trees are prone to overfitting. A random forest is an ensemble of decision trees, which is generated by randomizing features made available at every tree node. The authors chose for this method as it does not require hyper-parameter tuning, like methods as Extreme Gradient Boosting require. A discussion of the Extreme Gradient Boosting algorithm is given in Section 6.4.1, and this parameter estimation is given in 6.5.4. They show that the machine learning method performed better than the statistical approach in terms of prediction accuracy and computation time. The features that were most important are price, days to departure and schedule.

The rise of social media meant airlines have more ways than ever to communicate to their customers. Gunarathne et al. [62] show the effects of social media on the service levels of the industry. They use social media data, in particular Twitter, to model customer choice behavior. To accomplish this, they follow a bag-of-words approach to group similar tweets into clusters using a K-means algorithm. One of the challenges of text mining over Twitter messages is the high dimensionality of text. To get around this, latent semantic analysis (LSA) techniques were used to reduce dimensionality. Each Twitter message is represented in each own vector. LSA works by calculating the dot product between the normalizations of the two vectors. Vectors with LSA values close to 1 are considered "similar" Twitter messages, and are grouped subsequently. This is how dimensionality is reduced. They find that airlines respond faster to those that have more followers. Next, they cluster customers in five personality traits: openness, conscientiousness, extraversion, agreeableness and neuroticism. These personality traits are then included in the model, which also contains hashtags used, offensive language, competing airline mentioned and number of followers. They find that the number of followers, offensive language used and mention of a competing airline positively influence an airline's response time to an enquiry.

2.3 DEMAND UNCENSORING

Despite the large potential impact on profits, reviews of the RM literature show that the unconstraining problem has not received as much attention as perhaps it deserves, according to, for example, Bobb and Veral [63], McGill and Van Ryzin [64] and Talluri and Van Ryzin [65]. Indeed, research on unconstraining only meaningfully began in the mid-1990s. For some airlines, whose revenue is dominated by the income from more expensive fare-classes which are rarely constrained, this is perhaps more understandable. They may not feel that the benefits of unconstraining outweigh the costs of incorporating and maintaining a system for unconstraining within their existing complex computer systems. However, for many major airlines, unconstraining is important for both revenue optimisation and route/expansion planning.

Consider, for example, revenue optimisation for legacy airlines, which largely use capacity allocation rather than dynamic pricing. The estimation of true historical demand at different price levels informs these airlines' attempts to set optimal prices for the various fare classes on offer. Moreover, when these airlines offer multi-leg flights across a large network, different flight routes regularly share one or more legs. A crucial step in revenue optimisation is therefore deciding how many seats on each leg to allocate to each flight route, an optimisation problem that is usually formulated as an approximate dynamic program, as found in Talluri and Van Ryzin [65]. This formulation essentially says that the airline wants to allocate seats in a way which maximises total revenue received, without overbooking any individual flight leg, and without allocating more seats to a specific fare class than there is demand in order to avoid empty seats and lost revenue. Therefore, an accurate estimate of the *true demand* over the current booking window for each fare class is a crucial component of the revenue optimisation process. Inaccurate demand estimates will lead to poor seat allocations and thereby lost revenue, which highlights the importance of unconstraining.

In what follows, we categorise and give a brief overview of the main single-class unconstraining methods that have been proposed. For more general reviews of existing unconstraining research, see the work by Guo et al. [66] or Weatherford [67].

The most rudimentary approaches to dealing with constrained data involve no mathematics at all. One such approach is to simply ignore the fact that the data is constrained, and another is to disregard the constrained data entirely, basing forecasts exclusively on true historical demand data. The former approach, sometimes referred to as Naïve 1 or N1, as found in Guo et al. [66], will of course lead to (possibly very large) underestimation of current and future demand, which can potentially cause a 'spiral-down' in total revenue. This is illustrated by Cooper et al. [68] and Guo et al. [66]. The latter approach, sometimes referred to as Naïve 2 or N2, by for example, Guo et al. [66], may perform well in particular circumstances, for example, when only a very small number of data points are constrained, but in practice, this method can produce both significant over- and under-estimations of demand, depending on the context. This is illustrated by Zeni [69].

All other methods in the literature that deal with constrained demand data employ a mathematical or statistical model for the purposes of unconstraining. We can divide the remaining methods into two categories based on an important conceptual difference in their approaches to unconstraining. Methods in the first category (which we term '**multi-curve methods**') are applied to a set of historical demand data from a group of past flights. The historical order in which these flights occurred is irrelevant, and the goal of the methods is to produce unconstrained estimates for the constrained elements of that data set. The second category (which we term '**single-curve methods**'), consists of methods which are applied to one constrained demand curve at a time. Single-curve methods use demand data from a given flight up until the time that flight was constrained, to extrapolate what the true demand for that flight would have been, had it not been constrained.

The vast majority of existing unconstraining methods are multi-curve methods. One of the most elementary approaches in this category is known as mean-imputation, (or alternatively as Naïve 3, N3, or the 'mixed approach'). This method involves simply comparing each constrained value with the mean of all unconstrained values, and replacing it with the larger of the two. This is described by Zeni [69]. Variations of this method use the median or some other specified percentile instead of the mean.

Salch [70] proposed using the general statistical method of Expectation Maximisation (EM) for single class unconstraining in airline industry problems, and since then it has established itself as perhaps the most widely used unconstraining method. A variant of EM known as Projection Detruncation (PD) was first proposed for the purposes of unconstraining demand by Hopperstad [71]. Out of EM and PD, only EM has a rigorous statistical basis and has been proven to converge (under suitable assumptions [72]), as PD is based on a heuristic.

Skwarek [73] proposed a method known as Pickup Detruncation, which calculates the amount of bookings made for a given flight over the period it was constrained as the average total bookings made in the same period before departure for flights that were not constrained. Around the same time, Wickham introduced the Booking Profile (BP) method [74] (alternatively known as the 'multiplicative method'). BP works by using historical (true) bookings data from different flights to build a bookings profile over time, from the day tickets go on sale until departure.

Van Ryzin and McGill [75] applied the statistical method of Life Tables (LT) to unconstraining demand. Unfortunately, the method tends to produce biased estimates [76], and only produces unconstrained approximations of the mean and standard deviation, rather than estimates for each instance in which demand was constrained. Liu et al. [77] proposed a method for use in the hospitality industry which uses parametric regression. It differs most notably from other methods in that it attempts to account for other demand-influencing factors when calculating the distribution of demand, such as length of hotel stay and competitors' room rates.

The only existing method which falls decidedly into the 'single-curve methods' category was proposed by Queenan et al. [78]. They propose using the established forecasting algorithm known as Double Exponential Smoothing (DES), or "Holt's Method", for unconstraining demand, and compare its performance to EM, PD, LT and a variant of mean imputation. They report that, while in some cases EM outperforms DES, DES generally performs better on the most common booking curves shapes, and when the vast majority of the data is constrained. We discuss this paper in depth in Section 5.4.1.

Prior to Queenan et al. [78], a number of other papers had been published comparing the accuracy and revenue impact of many of the methods described above. Guo at al. [66] report that [79] and [80] compare the revenue impact of N2, N3, BP, and PD, finding that BP and PD outperform the Naïve methods. Weatherford [81] finds that out of EM, BP, N1, N2, and N3, the EM method best minimises the mean absolute error and best approximates the true mean of the data. Weatherford and Polt [82] and Zeni [69] compare EM, BP, N1, N2, N3, and PD, concluding that EM and PD are the best performing methods. The primary take-home message from these comparisons is that EM, PD, and DES are the most competitive and widely-used single-class unconstraining methods developed so far.

2.4 Cancellation prediction

Cancellation research can be divided into two different categories: methods using traditional, statistical methods, and those found using new techniques, through data mining and machine learning. Most of the work in the field of forecasting focuses on demand, not cancellations. There is substantial work in the field of revenue optimization that includes cancellations, but in these cases it is assumed the cancellation probabilities are given. In this literature review, estimating these cancellation probabilities is discussed. Statistical methods for cancellations forecasting are typically based on time series techniques such as (S)ARIMA. In the following sections, we will make the distinction between models that offer a probability on a reservation (PNR) level, and those that model based on aggregated level (such as booking class). In Section 2.4.1, we will provide an overview of forecasting cancellations on an aggregated level. In Section 2.4.2 we cover the non-aggregated, PNR, level approach. Note that both of these approaches have up- and downsides: using aggregated data may reduce the variance of number of cancellations, but at the cost of losing valuable information of individual cancellations. By using a PNR approach, all information is used but high levels of dimensionality and large variances may make predictions difficult.

Talluri and Van Ryzin [65], in their book, provide a great overview of cancellations and no-shows, but tend to focus on the optimization. While they introduce the aforementioned difference between an aggregated and non-aggregated approach, the actual modelling of cancellations is limited and assumed to be given. Instead, they focus on "net" demand. Net demand is the difference between all bookings and all cancellations. For an overview of how net demand may be calculated, please refer to Section 4.2.4.

2.4.1 *Aggregated-level*

The first known research that covers cancellations to my knowledge was the work from Beckmann and Bobkoski [83]. Until this point, cancellations were not considered in research. Beckmann and Bobkoski make the statement that to calculate a more accurate way of the expected value of revenue, is to explicitly introduce probabilities of cancellations, no-shows, misconnections and standby demand. They claim that these should be considered separately. In their work, they investigate distributions of demand, using nine months of data from one unnamed airline. While they point out a limited number of data points, histograms seem to indicate that cancellations exhibit an Exponential distribution. We would like to stress that this work is from 1958, long before the deregulation of airlines in the US, so these results may not be applicable today. Other results, that include analysis on demand booking curves and standby are shown, which is outside the scope of this section.

Martinez and Sanchez [84] followed by using data from a Spanish airline. They used reservation and cancellation data and use convolution of probability distributions based on this empirical data to obtain estimates for cancellation probabilities. Interestingly, they introduce the "forgetfulness law", and find that for international reservations, the probability of cancellation is independent of the date the reservation was originally made. This work, from 1970, is in contract with my findings, which is shown in Section 7.3.1. This seems to indicate that customer behavior has changed over the past fifty years.

It is often assumed that combining multiple models, an "ensemble" improves forecasting performance. For example, random forecasts most often outperform a single decision tree. Lemke et al. [85] explore the possibility of using different models to predict cancellations. Lemke et al. claim that (S)ARIMA models cannot be used for cancellations, since the time series themselves are very short in nature. They also note that computation time is important in practice, and therefore do not use time series methods. Therefore, they combine three models: single exponential smoothing, Brown's exponential smoothing and a regression approach. They then propose five different methods to combine these individual forecasts. They call these simple average (straight average), simple average with trimming (only the best 80% were taken), outperformance model (individual forecasts are weighted based on past performance), variance-based model (individual forecasts weighted based on past error variance) and optimal model (using covariance information). Interestingly, *none* of these methods generate a more accurate forecast. On the contrary, performance is worse than an individual forecast. They argue that forecasts on different level of aggregations may result in better performance. A similar thought is introducing additional information into a forecast. After all, intuitively it makes sense to make a forecaster method aware of external factors. This is introduced in Chapter 7.

Petraru [86] proposes an aggregate method for cancellations. Rather than building models on PNR level, Petraru discusses four different cancellation methods on an aggregated, booking class level. The first method only uses the number of bookings and associated cancellations made within that time frame. The second method separates bookings in hand and bookings to come, and it is assumed airlines know the number of cancellations in a time frame. In his third method, airlines know both the number of bookings and cancellations by time frame. His fourth method is similar to the third method but are scaled up and down for overbooking. All of the methods proposed in his thesis differ in the type of data used; there is no modelling in this work. Note that methods two through four simply use more data. Through simulation, revenue gains in the range of 1 to 3% are reported over the first method.

Cancellations in the hospitality industry were discussed by Liu [87]. He shows that for hotels, cancellation rates can be up to 60%. After unconstraining demand, he shows that for smaller hotels, a Poisson distribution is a good fit for demand distribution while the Normal distribution is a better fit for larger hotels. This is to be expected, since a $Poi(\lambda)$ approximates a Normal distribution as λ grows large. He shows that the type of hotel (airport, resort, city) has great effect on these rates. He also finds that the day of week has no different patterns in cancellation percentages. He proposes logistic regression to model cancellation rates.

2.4.2 *PNR-level*

Research into PNR-level cancellation forecast is relatively new. After all, disaggregated data has, by definition, a large number of dimensions. Machine learning methods such as random forests require large computational power. In this section, we will explore cancellation research in the framework of RM for the aviation, hospitality, restaurant and medical industry.

The work by Hueglin and Vannotti [88] was the first to my knowledge. They used data for bookings made over the course of a year for departure dates between one day and one year in the future. This is important since this ensures there are enough observations for different times in the booking curve. Attributes available to them were time before departure, booking class, class of service, number of passengers in the booking, origin and destination and weekday of departure. They then engage in feature engineering and generate features such as whether the PNR was split during its history, regions of departure, number of segments in the PNR, position of each flight in the PNR, flight time, connection time, purpose of travel and number of scheduled of flights per week. This last attribute is arguably used as a proxy for the presence of competition in this market. They deploy classification trees and logistic regression models on a PNR level, then aggregate these numbers to obtain the expected number of cancellations on a flight-level. They show that this method outperforms model that use aggregated (such as on flight-level) cancellations as input.

A great overview of cancellations in the hospitality industry can be found by De Korte [89]. De Korte briefly discusses cancellations in the hospitality industry and the consequences of errors in cancellation estimates. An important statement made is that cancellation rates are typically calculated assuming the absence of any competitors, that is, behavior is driven solely by the hotel itself; its competitors have no effect. He also reports that cancellation rates between 20% and 30% are not uncommon in the hospitality industry.

Morales and Wang [90] investigate cancellations for the hospitality industry. Working together with a major hotel chain in the United Kingdom, they collected 240.000 reservations, for check in dates between 2004 and 2006. The variables that were available to them is time of booking, whether a reservation is refundable, company on file, product, market sector, agent, channel, system, length of stay, room type, price, date of service and whether a reservation is a group booking. In order to gain an understanding of each of these features, they look at information gain. Interestingly, the time of booking was found to have the highest information gain. This is contrary to what Martinez and Sanchez [84] found, which we covered in Section 2.4.1, and is in line with my findings, which we cover in Chapter 7. Morales and Wang then compare logistic regression with support vector machines. First, just like Hueglin and Vannotti [88], they conclude that PNR models are more accurate than aggregated models. They show that while logistic regression shows promise, support vector machines perform better and reduce the error compared to these aggregated models by about 30%. They also report lower errors for reservations made closer to check-in dates. This seems to indicate that reservations made far in advance are more difficult to predict. Intuitively this makes sense, as there simply is more time for a customer to change their mind and cancel. They also conclude that at different stages in the booking curve, aspects differ in weights, and they therefore suggest building different models for different stages in the booking curve. We reached the same conclusion in Chapter 7.

This work of Morales and Wang [90] was extended by Antonio et al. [91] They use real data from four hotels (resorts) and work directly with the database used by these hotels. This is important because this shows how these methods can be used by other hotels in the industry. Additional features they acquired are, among others, the number of previous stays and cancellations. This brings an additional angle to PNR forecasting, making it more personalized. They discuss five different models: a boosted decision tree, decision forest (also known as a random forest), decision jungle (an extension of a decision forest), support vector machine and neural network. They find that out of these models, a decision forest works better. Neural networks seem to consistently perform worst, but we would like to stress that neural networks require more user inputs than a decision forecast, so this may not be a fair comparison. In Antonio's book [92] this work is repeated and shown how his approach works in different hotels.

As discussed in Section 2.4.1, intuitively it makes sense to make the forecasting model aware of external effects. See, for example, Talluri and van Ryzin [65]. This is further explored by Antonio and Nunes [93]. They revisit their forecasting model from Antonio et al. [91] by including online reviews from TripAdvisor.com and Booking.com. The former was briefly introduced in Section 2.2.1. It allows users to post reviews and photos of hotels they have stayed at. Booking.com is the global leader in the online travel agent domain. Customers can not only book hotel rooms, but also leave reviews here. Both of these sources were included. Moreover, pricing and inventory available from Booking.com were also included. Calendars with events, school holidays and the weather forecast include variables that were included in their models. They find that incorporating this additional information did *not* improve the model's performance. Tsai [94] does show improved results by introducing temporal features in his model to predict cancellations. By using features such as how each feature has been performing recently and reliability of schedules over time, he shows improvements over traditional regression models.

The restaurant industry is investigated by Huang et al [95]. Restaurants, just like airlines and hotels, face having limited capacity and a perishable asset. Most restaurants let patrons reserve a table without fees. This, in turn, means that the no-show percentage is much higher than in airlines and hotels, that do often have a no-show penalty. These no-shows expose the restaurants to risk: these tables could have been assigned to others, or waiting walk-ins. Therefore, predicting cancellations for restaurants is just as important as it is for airlines or hotels. They accomplish this by building a neural network based on date, whether it was a holiday, gender, age, income, education level, marital status, place of residence, cancellation record, and cumulative number of cancellations (if present). While research by Antonio et al. [91] showed that neural networks do not perform as well as decision forecasts, Huang et al. [95] show promising results. They compare a back propagation neural network, the most widely used neural network, to a general regression neural network. They discuss benefits of this approach, but the main benefit is its increased learning ability. Interestingly, overall model performance is comparable. The aforementioned method is better at predicting true positives, while the latter is better at predicting true negatives. Predicting true negatives is found to be the hardest.

Cancellations in the medical industry are explored by Alaeddini et al. [96] They discuss both cancellations and no-shows. These disruptions not only cause inconvenience to doctors, nurses and hospital management alike, they also have a direct impact on revenue, cost and resource utilization in healthcare systems. This causes other patients not being able to get timely appointments because part of the schedule is filled with patients who will not cancel or no-show. Also, when scheduled patients cancel their appointments, they often leave the clinic with a very short amount of time to fill the schedule. Alaeddini et al. discuss up- and downsides of aggregated and disaggregated models. They then propose a hybrid model: first, a multinomial logistic regression model is built to estimate no-show and cancellation probabilities. This is done using aggregated data. Next, Bayesian inference is used to personalize these estimates on an individual level. Contrary to Lemke et al. [85], discussed in Section 2.4.1, very promising results are shown by combining forecasts. Methods are compared to twelve other methods in the literature, and all of these are outperformed.

2.5 PRICING

In this section, we provide an overview of the current theory of pricing. The main objective of pricing in airlines is determining the "optimal" fare levels it offers to customers. Finding these "optimal" fares ultimately depend on the airline's objective (these can be, for example, maximize profit, maximize revenue, maximize market share). This topic consists of multiple parts: segmenting the market (this is achieved through different tools, which we discuss in Section 2.5.1), setting the actual prices (introduced in Section 2.5.3) and customer perception (discussed in Section 2.5.3.

Talluri and Van Ryzin [65] provide an excellent framework on the theory and practice of RM. They make the differentiation between quantity-based RM and pricing-based RM. They provide several models for pricing-based RM, and provide algorithms to set optimal pricing in the deterministic and stochastic cases.

2.5.1 *Segmentation*

Botimer [97] was the first to provide a detailed brief of airline pricing and product differentiation. In his PhD thesis, he provides an overview of traditional tools for segmentation. His first definition, first degree price segmentation, assumes that an airline is perfectly able to identify and segment each and every potential customer. Realizing that this is unrealistic, a second degree of price discrimination is introduced. In this instance, airlines estimating individual's willingness to pay through their historic purchasing behavior. A third degree of discrimination is introduced: in this case, the airline segments the market a-priori based on self-imposed segments (for example, students under 25 years old), without looking at its data. It is identified that airlines use second degree price discrimination. Botimer follows by introducing the most common tools used: black out periods, which makes it impossible for customers to purchase a (low) fare. A Saturday night minimum stay, which calls for customers to spend at least a Saturday night at their destination before returning home. Or, forcing a customer to buy a roundtrip for a given product. Botimer's thesis, while almost 30 years old now, provides a good background on segmentation by airlines.

Fare fencing is the process of ensuring that products are segmented in such a way that customers are not able to purchase lower-priced products than the airline intends to. A classic example of fare fencing is the usage of forcing a Saturday-night stay before returning home for cheap fares. It is typically thought that business travellers prefer to return from meetings before Friday, so they spend the weekend at home. It is thought that leisure travellers do not mind spending a Saturday night at their destination. Therefore, the "Saturday-night" rule is thought of being an effective way to "fence" products. Business travellers are unlikely to "cross" this fence and purchase the lower fare. This is discussed by Zhang and Bell [98]. After identifying segmentation tools that align with Botimer [97], they discuss business related issues to segmentation and discuss a bird's eye view of the practical implications of fare fencing. Many of these issues showcase the inability to fence products: for example, it is very difficult to fence different products by age of the customer through traditional distribution channels.

The importance of customer thinking is discussed by Shen and Su [99]. They provide an overview of customer behavior modelling. In their paper, they discuss the concept of a strategic customer. As they state, it is typically assumed that once a customer arrives, a product is purchased, or no purchase is made and this customer is lost forever. They claim that the industry has forgotten about customers that delay their decision making. We cover this particular topic in Chapter 8. Next, they introduce the concept of dynamic pricing. It is claimed that if firms fail to account for strategic customer behavior, the loss in revenue can be substantial, reinforcing the importance of my work in Chapter 8.

Vinod [100] discuss the evolution of the customer in RM. He claims that a customer goes through five different stages: an unqualified customer becomes a prospective customer through options available in the market. Once a ticket is sold to this individual, this individual becomes a customer. This customer transforms into a satisfied customer once perceived quality of service is fulfilled. Finally, once this satisfied customer is cared for, it becomes a repeat customer. According to Vinod, the optimal way to segment customers is done through operational data (PNR data), customer profile (name, address), demographic data (age, income) and preferences (behaviors). He then notes that segmenting customers through fare rules, which were introduced by Botimer [97], is only one-fourth of the optimal way of segmentation. He follows by stating that the traditional way of selling tickets, through different distributional channels, make it impossible to truly segment customers effectively.

2.5.2 *Setting fares*

In this section, we will review the literature on setting fares. In practice, fares are most often set looking at competition. There are many reasons for this, that are inherent to RM and to human behavior: risk-adverseness. Not one airline would like to start trying something new, in the fear of losing demand. Once a flight takes off, any empty seat could have been sold through a better pricing structure. This inherent fear inhibits pricing managers make decisions that pose a larger degree of risk than decisions that are deemed to have a lower degree of risk.

Garrow et al. [101] provide an overview of pricing in different industries, including the airline industry, at the time (2006). They first identify centralisation of pricing decisions and alignment with sales compensation. Next, they identify the concept of pricing the experience, which is often seen in practice. Next, bundling is identified as an important topic in setting fares, but this is something that is more relevant to travel agent than airlines. The impact of distribution channels is discussed, next. Most airlines engage in full-content agreements (FCA), which is a contract that says that fares offered are identical regardless of distribution channel. The difficulties and tensions that ensuring this is followed in practice are exposed.

The process that airlines follow after fare values have been decided, is discussed by Poelt in Yeoman's book [102]. He gives an overview of the pricing data flow in practice and illustrates the role of ATPCo (Airline Tariff Publishing Company, the company that airlines file their fares to), GDS (the instrument that ensures that fares are available for the different distribution channels), sales and airline's RM systems.

Bitran [103] provides a stochastic control problem that dynamically sets the products and associated prices for sale so as to maximize total revenue. Bitran's work starts with a single-product case, assuming deterministic demand. Given an initial inventory, a single product, he formulates a stochastic problem that calculates an optimal policy. This method is extended by allowing multiple products for sale. Next, he introduces stochastic models for the single and multiple product case. Bitran compares the outputs with the work of Belobaba [13] and notes performance gains. While Bitran's work shows potential, we feel it is important to stress that he assumes that a price sensitivity curve is given. Determining a true price sensitivity curve in practice is extremely complicated. Also, he assumes an airline has a monopoly: there are no competing airlines.

Boyd [104] discuss this first shortcoming. In practice, it is difficult to establish a customer's willingness to pay. Boyd hypothesizes the effects on customer behavior when customer only purchase the lowest fare. Hopman et al. [22] provide an optimization technique for this behavior. They assume that every customer purchases the lowest available fare for sale. Significant revenue gains are shown. This is extended by introducing an optimisation method in which the fare offered to the customer may never decrease. This avoids strategic customer behavior, which we discussed in Section 2.5.1. They show negative revenue (2%) performance if there are no customers that postpone their decision. However, once the number of customers that postpone their decision is increased, significant revenue gains are reported.

Vinod et al. [105] discuss this second shortcoming. In their framework, they aim to calculate the attractiveness of a given itinerary. This is based on travel time, time of day, relative fare, and presence of an airline in the airport of origin. This is modelled using a MNL model. They then propose their framework: after establishing an initial estimate using the MNL model, it is iteratively updated by incorporating competitor airlines until some set threshold is reached. Having established the attractiveness of the airline's option, and its competition, they formulate a non-linear programming model. It is then discussed how this output can be fed directly to distribution channels, avoiding the intervention of a pricing professional in the RM department. It was found, through simulation studies, that the most expensive products would benefit from slightly lower fares, and the cheapest products would show gains from slightly increasing their price.

A fare is only available for sale if the airline offers inventory for that particular fare. For example, consider a product priced at $1, by mistake. If the airline does not make this product available for sale, then customers are not able to purchase this product. Therefore, it is typically thought that controlling products through inventory can solve suboptimal pricing levels. This process, called fare management, is discussed by Vinod [106]. He investigates the importance of setting the right fare. He illustrates the complexities of setting the right fares, and show how ineffective fares lead to losses through errors in suboptimal control.

Related to this, is the concept of fare adjustment. In order to determine what product is available for sale, its fare value is compared to an opportunity cost. Traditionally, this fare value that is compared is, simply, the price of the product. A fare adjustment is a positive or negative change to the price of a product. A positive fare adjustment makes it more likely that the opportunity cost is exceeded (and thus available for sale), a negative fare adjustment makes this harder. Westermann and Lancaster [107] claim that RM and distribution departments have seen improvements, but traditional pricing has received little focus. As a result, they argue, airlines should focus on developing pricing decision support tools, and actively improve the integration between RM and pricing. They highlight one particular example of integration: fare adjustment in terms of buy-up and buy-down behavior. By using a fare adjustment, pricing and RM teams are aligned: the effects of pricing are embedded into the optimisation aspect of RM. One of these decision support tools for pricing is introduced by Ratliff and Vinod [108]. They draw parallels with the revenue opportunity model, and design a similar model for pricing. A revenue opportunity model is a method of optimization that calculates optimal revenue for historical departures. Since, in hindsight, there is no such thing as stochastic demand, often a linear program with deterministic demand is used. The actual values of number of products sold are then compared with the optimal value to understand what products were over- and under-sold, as a result of a suboptimal policy that was calculated. Ratliff and Vinod introduces a workflow that supports proactive pricing (planned prices aligned with sales and marketing departments) and price leadership (be the driver, not a follower). They propose revenue is the multiplication of price, market share and market size. Through the use of airline's own and competitor fares, empirical match factors, historical ticket sales, taxes, market shares and price elasticities, optimal price points are calculated. Through this model, they aim to minimize pricing manager intervention of setting fare levels, which is often done without analysis but rather done by trial-and-error.

Another concept in pricing is ancillaries. The pricing of ancillaries has been of great interest to the industry. Ancillaries are all products outside the ticket a customer purchases. An example of an ancillary is additional luggage offered to the customer. Parker [109] discuss ancillaries, fare families and a-la-carte pricing. He acknowledges that airlines have typically sold fares from the bottom up, from the lowest qualified fare. However, ancillaries are making it possible for airlines to sell from the middle or top, based on individual preferences. Parker sets out examples of ancillaries, such as the usage of mobile phones, onboard internet access, eyeshades and socks, onboard pillow and blanket, in-flight entertainment and handheld entertainment devices. One of his most important findings is, similar to Choi and Mattila [110]: airlines should make privileges associated with different fare products clear to the customer.

The financial risk of RM is discussed by Lancaster [111]. Meltdowns of the airline industry are discussed and the effects and the risks these pose on RM systems are shown. Lancaster discusses sources of risk, suggest ways of quantifying this risk, and how to compare between different methods that have different levels of risk. Lancaster introduces a financial approach to RM, by introducing the concept of Value-at-Risk (VaR). This financial measure estimates how much a financial product, in this case, airline revenue, may lose as a function of probability. Lancaster suggests taking the RASM (the total revenue divided by the sum of the multiplication of seats and distance) and to reformulate this to add a measure of risk.

Tretheway [112] discusses the modus operandi of full service and low cost airlines. He claims that the model of legacy airlines is broken, and puts forth several arguments based on how lost cost airlines have changed the field of pricing. He identifies the introduction of one-way pricing for low cost airlines. Low cost airlines often have one-way fares that make up a return fare, while legacy airlines have specific (more expensive) one-way and return fares. Next, the loss of true segmentation ability is identified. We have identified this above: airlines are not able to segment their markets are they would like. Next, he claims that legacy carriers, often with a greater network, misuse the concept of "beyond revenue". For itineraries consisting of more than one flight, how is the total revenue distributed among flights? In other words, what is the "beyond revenue" for the second flight? It is claimed that many airlines overestimate this and use this to justify their network structure and price levels. Lastly, he claims that pricing is only done short-term, while long-term pricing is often neglected. This is contrary what those in field of Economics suggest is necessary for long-term profitability.

The hospitality industry was further explored by Ivanov [113]. This work provides an excellent framework of RM in this industry. The RM system, process and metrics are discussed. This is followed by market segmentation. While airlines traditionally segment customers into two groups: leisure and business travellers, the hospitality industry segments the market into fully independent tourists (individuals), leisure tourists (groups), senior travellers, business travellers (trips, special events), families with children and flight crew. Next, the concept of perceived value is discussed. This is separated between tangible attributes (for example, hotel facilities) and intangible attributes (for example, speed of service). Examples of these are given and it is shown how hotels position themselves. When it comes to pricing specifically, pricing managers face three dimensions: price length (price varying depending on time), fare width (price consistency) and fare depth (prices for different distributors). While the author does not go into detail how prices should be set, Ivanov provides a great framework for pricing managers. Pricing in the hospitality industry is discussed in detail by Noone and Mattila [114]. In their work, they study not only setting the actual pricing, but also how these prices should be shown to customers. They introduce the concepts of blended and non-blended rates. A blended rate is the average nightly rate, while a non-blended rate is a list of (the actual) nightly rates. They find that showing rates for individual nights is more effective than showing the sum of rates for all nights. An increase in willingness to pay is reported. Noone and Mattila also report that it is important for hotels to make the rates *understandable* to customers. They stress, for example, that if a customer benefits from a lower nightly rate by booking multiple nights, this should be communicated and explained to the customer. They find that this, as well, increases willingness to pay.

Other industries are discussed by Cross et al. [115] He discusses early day RM in the airline industry (American Airlines, selling "early bird" tickets), the early hospitality industry (Marriott, optimizing rate offered by length of stay), courier industry (United Postal Service, target pricing: what is the optimal price level to offer to customers), automotive industry (Ford, different price based on geography, type of buyer and product configuration) and recent advances in the hospitality industry (Intercontinental Hotel Group, increasing revenue per available room by 2.7%). Bodea and Ferguson [116] introduce a framework how to segment markets and set optimal prices in different industries. In Chapter 6 specifically, they discuss the theory and practice of pricing. In the theory section, they introduce price elasticity. Different industries are compared. Next, price response curves are discussed. The maturity of pricing expertise in any company can be measured by pricing expertise and data availability. Next, they suggest how price elasticity is to be modelled. In the next chapter, dynamic pricing and markdowns are discussed. The authors shed light on why permanent price markdowns are so often used and claim that this is often detrimental to an organization. Instead, an optimal price markdown policy is given, with the explicit goal to maximize gross margins.

2.5.3 *Customer perception*

While it is vital for airlines to set fares, a customer's perception of its pricing strategies are just as, if not more important. In this Section, we will explore how customers perceive pricing.

The hospitality industry is discussed by Choi and Mattila [117]. They conducted an experiment in an airport, posing travellers questions. First, they distinguish between fixed (hotel rates are the same year around) and variable (hotel rates vary based on seasonality) pricing. Next, the respondents were given information on the rate others were given in the same time period. Lastly, half of the respondents were told about the practice of RM, while the others were kept in the dark. They then study what the effects are on the perceived fairness of RM. Naturally, they find that respondents that were given higher price points than others (and knew they were offered a higher rate) perceived RM less fair than those that were offered lower fares. When customers compared their rate to others, and the found their rate to be higher, this was perceived to be unfair. However, the opposite was not found to be true: when a rate was lower, this was *not* perceived as being unfair. This work is continued in Choi and Mattila [110]. In this study, they claim that customers have become aware of RM tactics, and claim that revenue managers have started to worry whether customers deem their practice fair. To investigate, they design three scenarios: no information (room rate is offered, but no explanation), limited information (room rate dependent on day of week, length of stay and time before check-in date booked) and full information (all of the above, plus pricing insights to the customer: weekend stays cheaper than week days, stays booked far in advance cheaper). They then conduct a similar survey as in their earlier work [117]. Perceived fairness improves by 8% when using limited information, compared to the no information scenario. However, the biggest improvement can be seen by introducing full information: perceived fairness improves by 20%.

The restaurant industry is investigated by Wirtz and Kimes [118]. First, they give background on perceived fairness in general. They introduce the definition of dual entitlement: customers believe they are entitled to a reasonable price, while believing that a firm is entitled to a reasonable profit. Furthermore, an increase in price is considered to be fair by customers if it comes at an increased cost to the company. Since the level of service offered by airlines are fundamentally the same (a seat is a seat) without additional cost, they claim that customers may perceive RM as unfair: increases in price at no (additional) cost to the company. They find that customers that *benefit* from discounted fares consider RM fair, while those that don't have access to products, as a result of fare fencing, find themselves disadvantaged. They conduct two studies in a restaurant and in a hotel. They find that, just like Choi and Mattila [117] found in the hospitality industry, that customers that are familiar with the underlying factors that affect pricing, are much more likely to perceive pricing as fair. They conclude that not just the process of variable pricing is to be communicated to customers, but also to create familiarity with those practices. They specifically find that customers that are aware of this type of fencing, do not share this negative perception. Kimes and Wirtz [119] discuss the fairness of restaurant RM through surveys across three different countries. They study the concept of a reference price. This is a "suggested" price that customers become accustomed to seeing when making a purchase. Interestingly, they find that pricing that was originally seen as unfair, such as hotel rates, have become accepted by society. They also illustrate how cultural differences affect the acceptance of variable pricing and perceived fairness. In a survey, they study price discrimination in terms of time of day (between lunch and dinner), day of week (weekday or weekend), time of arrival (early or late dinner), table location (away or close to the window) and coupons (two-for-one). It was found that time of day, coupon and time of arrival were deemed fair. Varying pricing based on day of week was considered neutral. The perceived fairness of pricing based on table location was found to be unfair.

The golf industry was studied by Kimes and Wirtz [120]. They find that many of the properties of RM (in particular, fixed capacity, perishable goods) can be found in the golf industry, but there has been little evidence that RM strategies are used in practice. One particular reason for this, they hypothesize, is customer backlash. They too introduce the concept of a "suggested", or "reference" price. These prices may come from market prices, posted prices and past experience with a company. For example, they claim that if a round of golf is priced around $75 in the market, the customer's perceived fair price is around $75. To study this in the golf industry, they conducted a survey with six scenarios: pricing based on time of day (peak or off-peak), varying price levels (same day, different rate), coupons (two for one), time of booking (discounting early bookings), reservation fees (waived if honored, no-show fee charged if customer does not show) and tee time interval pricing (price increases as time spent on the golf course increases). For each of these scenarios, they distinguish between a positive (typically lower pricing, a discount) or negative (typically higher pricing, a premium) change. Time of day pricing was perceived to be fair, but offering a discount was more acceptable than a premium. Varying pricing (this week on Saturday $50, next Saturday $40, week thereafter $60) was found to be extremely unfair, both discounts and premiums. Discounts through coupons were perceived to be fair. Time of booking pricing was deemed unfair, both for discounts and premiums. On the other hand, reservation fees, in particular charging a no-show fees for customers that do not show up, was deemed fair. Tee time-based pricing was perceived as fair, as well. Interestingly, some parts of pricing (no-show fees, time of day) used by the aviation industry are considered acceptable while others (varying pricing, time of booking) are not.

2.6 OPTIMISATION

The end goal of RM is to determine the optimal booking policy. In this case, an optimal policy is the policy that maximizes (network) revenue. The field of optimization can be separated by assumptions on whether cancellations are modelled or not; by single resource or network optimization; and, finally, by assumptions on demand: independent or customer choice. Cancellations can either be modelled directly (for example, by explicitly calculating cancellation probabilities and using these in an optimisation strategy) or indirectly (by removing cancellations from demand inputs, and assuming cancellations won't occur). Wang et al. [121] provide an overview of challenges and progress in the RM field. They conclude nine emerging themes in RM and eight managerial shifts. One of these shifts is the change in demand forecasting from historical data to using big data techniques. It is interesting to note, however, that they do not mention cancellations at any stage. Next, we make the distinction between leg-level control, and network optimization. Leg-level control are methods that optimize flights individually, network optimization are methods that optimize an airline's network all at once. In a study conducted by Weatherford [122], it was reported that 38% of the airlines that responded use a network RM system. Therefore, a significant number of airlines still use leg-level optimization techniques. Finally, demand can be modelled as independent demand or through customer choice. When independent demand is assumed, a customer will either purchase a product he is forecast to buy, even though other products may be available for sale. On the other hand, dependent demand, modelled through customer choice assumes a customer is given a set of options and may choose any, or none at all.

First, we will review leg-level optimisation in Section 2.6.1. This is followed by network optimisation in Section 2.6.2.

2.6.1 *Leg-level optimisation*

The first works of optimization in RM are credited to Beckmann [24] and Littlewood [25]. Beckmann [24] uses continuous demand and finds an optimal policy by calculating a series of integrals. Littlewood [25] approaches the optimisation approach as a newsvendor problem. The newsvendor model is a mathematical model in which a newsvendor needs to decide how many newspapers to purchase at the beginning of the day. If a newsvendor purchases too many units, these become worthless (after all, nobody wants to read yesterday's paper!). However, if a newsvendor does not purchase enough, he could have made more money by increasing his inventory level. The parallels with (airline) RM are clear: any seats unsold are worthless and being sold out could have meant selling more of higher-priced products. This approach, introduced by Littlewood [25], is extended by Belobaba [13]. Belobaba calls his approach the Expected Marginal Seat Revenue (EMSR), and introduces a method to allow more than two classes. Under conditions, it can be shown that this method is optimal for two booking classes, but requires a heuristic for more than two. Both Littlewood and Belobaba assume aggregated demand over the booking curve; that is, there is no element of time.

The first work that incorporated time was Lee and Hersh [123]. In this work, he describes a time-dependent model. Demand is assumed to be independent and follows a Poisson process, and no cancellations are assumed. A solution is found by solving a dynamic program, being solved backwards in time. In this dynamic program, the value function represents the revenue-to-go, given a tuple of capacity and time left until flight departure. This approach relies on boundary conditions: once a flight closes for sale, the value function is equal to zero, regardless of capacity. In addition, once capacity is exhausted, the value function is also equal to zero. This is true since this formulation does not assume any cancellations. This dynamic program was the basis of many publications to come, and is covered in Chapter 8 as *DPID*. Lautenbacher and Stidham [124], extends the work by Lee and Hersh [123] by studying the underlying Markov Chain. They show that the optimal policy, without the element of time, as put forth by Beckmann [24] and Littlewood [25], can be expressed as a similar policy as the work by Lee and Hersh [123]. In their work, where they assume there are no cancellations, no-show nor overbooking, they show that the value function by Lee and Hersh [123] is concave and as a result the optimal policy found by solving the dynamic program can be translated in an optimal booking policy in terms of a booking limit. A booking limit is the number of seats that are available for sale for a given product, which is the very output of Beckmann [24] and Littlewood [25].

The inclusion of cancellations was first studied by Subramanian et al. [125] In this work, they analyze a Markov decision process through dynamic programming with cancellations, no-shows and overbooking. They allow multiple fare classes. Just like Lautenbacher and Stidham [124], they exploit an equivalence with a problem from queuing theory to transform a multi-dimensional state space (after all, it needs to keep track of number of seats sold for every class) into a single-dimensional dynamic program, by assuming cancellation probabilities are independent of fare class. While this assumption does violate what is seen in practice, it makes this problem tractable. They conclude with three main findings: first, booking limits may not be monotonic in time before departure. Second, it may be optimal to accept a lower priced request than a higher priced request. This violates the concept of nesting: if an airline is willing to accept a product priced at $100, it should intuitively also be willing to accept a product priced at $150. We reach the same conclusion in Chapter 9 (but for a different reason). Thirdly, the optimal policy depends both on total capacity and remaining capacity. This is important since the work by Lee and Hersh [123] only depends on capacity remaining. Gosavi et al. [126] extend this work by lifting the assumption of having class-independent cancellation probabilities. They solve the dynamic program using reinforcement learning, and show how scalable this technique is. They then compare the results of this method to the EMSR method of Belobaba [13] and show revenue gains. Just as important is the robustness of this work. They show that underestimation of probability of cancellation results in greater loss than overestimation. Moreover, they find that overestimation of demand in lower fare classes results in greater fare losses than underestimation. This bias of underforecasting lower fare classes and overforecasting higher fare classes is in line with what we report in Chapter 8.

Boyd and Kallesen [104] move away from the independent demand assumption by segmenting passengers between priceable (passengers that book by fare) and yieldable (passengers that book by product). Forecasting yieldable demand can be thought of as forecasting independent demand, for priceable demand a new forecasting technique needed to be developed. Belobaba and Hopperstad [5] describe one approach to forecast priceable demand. They forecast demand for the lowest class, and then estimate sell up probabilities to higher classes. Simulation-based optimization was studied by van Ryzin and Vulcano [127] or Vulcano et al. [54]

The dangers of buydown, which in the long term cause 'spiral down', are shown in a study by Cooper et al. [128] They develop a mathematical model that defines when spiral down occurs. Work on efficient frontiers, such as that of Phillips [129] and Fiig et al. [130] show what the effects are of buydown.

Fiig et al. [131] study the optimization of mixed fare structures, and create the notion of fare adjustment. They construct a choice model and use a marginal revenue transformation to both demand and fares to obtain what they refer to as EMSRb-MR limits. They show how to transform any choice model in such a way that results in an equivalent independent demand model. Changing these fares is referred to as fare adjustment. This translates into tighter booking limits for the classes with low fares. They compare the results in the passenger origin destination simulator (PODS) to hybrid forecasting with and without this fare adjustment.

In the optimisation problems covered above, the main decision variable is some form of booking limit: the number of seats to sell for a certain product. These may be calculated through either time-independent, such as EMSR, or time-dependent model, such as a dynamic program. A completely different approach is introduced by Frenk et al. [132] Instead, they use the closing time of sale for each product as a decision variable. They show potential for this approach, but find that this model does generate policies that are relatively nonrobust against assumptions of cancellation times. They do report revenue gains compared to the EMSR method, but only when cancellations exhibit hyperexponential cancellation times. Their model does suffer from large state space when class-dependent cancellation rates are introduced.

Hopman et al. [23] (see also Chapter 9) provide a dynamic programming approach with a heuristic to solve this problem. After establishing cancellation probabilities, a framework of which is given in [21], they use a heuristic to maintain a single-dimensional state space. This is important, they argue, to minimize computation time. This topic is further discussed in Chapter 4. Using real airline data, they show an example with three products, priced at $1000, $750, $500, in which it is optimal to close the product priced at $750 for sale, while keeping the product with fare $500 open to purchase. This is the result of relatively high cancellation probability of the more expensive product. The fare that is used to obtain the optimal policy is a risk-adjusted fare, based on the risk of time of cancellation.

Lastly, we would like to stress it is important to realize what airlines are optimizing for. It may be argued that all airlines want to maximize revenue. However, finding the optimal solution may not always be the solution that revenue managers *want* to see. This concept is reviewed by Gonsch [133]. Through a survey, he finds that most revenue managers are risk-adverse. He finds that most algorithms assume uncertainty of demand, but very few consider uncertainty of fares (most often, fares are assumed to be constant over time). Instead, Gonsch [133] introduces several methods that concern risk-adverse RM and hypothesizes that these are important in practice. Hopman et al. in [17] and Chapter 3 discuss these problems from a true practical perspective. They not only discuss the risk element of optimisation, but also practical limitations of finding inputs required for optimisation, system limitations and cultural challenges that airlines face in RM. They conclude that while theoretically airlines aim to maximize revenue, methods that are simple and understandable to revenue analysts and managers (but perform worse) are often preferred.

2.6.2 *Network optimisation*

There have been attempts to make leg-level controls possible for network RM. One of these metrics is found using approximate dynamic programming. In this approach, the value function of the network RM problem is approximated by a sum of flight value functions. This approach is given in Chapter 4. In this section, we review study *true* network optimisation problems. While all of these optimisation methods require some sort of approximation - after all, they suffer from the curse of dimensionality - the methods that are followed were originally *designed* for network RM.

Bertimas and Popescu [134] introduce the network RM problem. They first show the Bellman equation, assuming no cancellations, which state space consists of a vector with remaining capacities of every flight. It is clear that this problem grows large very fast: suppose an airline operates just five flights, each with a capacity of 100 seats, the state space is of size 100^5. Realizing this, they relax the notion of time by aggregating demand (expected demand) over remaining time periods and introduce an integer programming formulation that solves for the optimal policy. An integer programming formulation is necessary as naturally a fraction of seats cannot be sold. They relax this assumption by obtaining a linear program. Next, they propose approximate dynamic programming algorithms. First, optimal policies are discussed: bid price control (which accepts requests for a product if they exceed a static opportunity cost at the next time stage) and certainty equivalent control (requests are accepted by calculating the aforementioned two different linear programming with different capacities). One drawback of bid price control, is that it uses shadow prices of a linear program and these may not be unique. This is solved with the certainty equivalent control, however at the expense of calculating this linear program at every request, which is computationally heavy. They find, however, that this control outperforms bid price control by 1%. This is an important and interesting result, since most airlines to this day still use bid price control. The suboptimality of of bid prices in certain cases can be found in Talluri and Van Ryzin's book [65] (p. 90). However, in practice the consensus seems to be that a bid price strategy works well. well. In particular, it is agreed that frequent reoptimisation is necessary. Reoptimisation is required as forecasts may change over time, and capacity may change slower or faster than expected. Pimental et al. [135] show that infrequent reoptimisation can result in revenue losses of 6% in the hospitality industry.

Different authors propose different methods to solve the network RM problem. Shibab et al. [136] choose a different technique to combat the curse of dimensionality. They introduce a dynamic program with a state space consisting of time remaining and number of booked by class. Next, they discuss Q-learning (a technique that approximates the contribution of taking a decision being in a state), neural networks and deep Q-learning (a novel combination of the two former models) to solve this dynamic program. By incorporating deep Q-learning, they obtain revenues within 7% of the optimal solution. While these results are promising, only cancellation rates of up to 20% were used. In practice, many markets exhibit higher cancellation rates. Yet another approach is taken by Dai et al. [137] They study the network RM problem with cancellations and no-shows. They introduce both methods for independent demand and the customer choice equivalent. They show that methods are optimal policies are intractable and propose a deterministic, continuous-time and continuous-state model to solve this. To accomplish this, they use a fluid solution. While the method was developed for networks, this method is tested using 20 small-sized single-leg problems so optimal policies can be calculated. It was found that their approach obtain revenues less than 1% away from optimal revenues. The small-scale of this experiment raises questions whether this approach will work for (large) networks.

The RM problem is also tackled from another angle, by means of "choice-based" RM. Talluri and Van Ryzin [138] were one of the first to introduce the notion of a "choice set", where customers are being offered a set of options. These may be products with different conditions, for example. They purchase an option in the choice set with a certain probability, the goal is to find those subsets offered that maximizes total revenue. Identifying these sets, dubbed efficient sets, is computationally complex. The dynamic program that follows is hurt by the curse of dimensionality, and subsequently there has been research to tackle this by approximating the solution by a (deterministic) linear program, see Liu and Van Ryzin [139], for example. There have been several other approaches. Zhang and Adelman [140] use a different approach and approximate the dynamic program using a weighted basis function. Kunnumkal and Topaloglu [141] propose another approximate dynamic programming approach. An approximate dynamic programming formulation for network RM under customer choice was introduced by Zhang [142], but lacks cancellations. Bront et al. [143] propose an alternative and use column generation. Sierag et al. [144] were the first ones to consider the RM problem including cancellations and customer choice. Sierag et al. [145] further analyzes the single-leg RM problem under customer choice. Cancellations are independent of the current time in the booking curve. They demonstrate that their formulation is very robust against unknown cancellation behavior, which has substantial practical benefit as cancellations may not always be modelled with statistical significance. .

The performance of choice-based RM is investigated by Carrier and Weatherford [146]. They use the passenger origin and destination simulator PODS (passenger origin and destination simulator, a large simulation model with real airline inputs) to study the effects of using multinomial logit models in a competitive environment where airlines compete for passengers. They show that optimizing using MNL models outperforms standard forecasting, but is outperformed hybrid forecasting and fare adjustment. For a great overview of PODS, we refer the reader to Carrier [147]. For an extensive review of dependent demand RM, we refer to the work of Weatherford and Ratliff [148]. They provide an overview for both non-choice and choice-based methods, for both forecasting and optimization.

The importance of explicitly modelling cancellations is studied by Petraru [86]. He provides heuristics to estimate cancellation policies and uses PODS to show how each of these methods perform. He shows revenue gains between 1% and 3% over methods that do not use these heuristics. Interestingly, he shows that leg-level cancellation rates generate slightly higher revenues than OD/path cancellation rates. As with most PODS studies, due to its complexity of different forecasters, optimization methods, denied boarding costs, it is difficult to pinpoint exactly what causes these revenue gains, however.

The context of overbooking and cancellations in the restaurant industry was reviewed by Tse and Poon [149]. He considers cancellations, overbooking and walk-ins in the restaurant industry. While the aviation and hospitality industry makes a clear distinction between cancellations (a customer cancelling their reservation before a deadline, such as check-in date or closure of check-in desk) and no-shows (a customer not cancelling but simply not showing up), Tse and Poon suggest that these can be combined for the restaurant industry. For more background of cancellations in the aviation industry, please refer to Chapter 7. Using data from a restaurant in a hotel in Hong Kong, they propose to model cancellations using a Binomial distribution. Interestingly, they report low cancellation rates, around 10%. They also report that for this restaurant, there are statistically significant difference between cancellation rates by day of the week. One key difference the authors point out between the restaurant and hotel industry that units of capacity (tables) may be used fractionally. For example, a table that seats six may be seated by a couple, if deemed necessary. They also highlight the importance of the link between cancellations and price sensitivity. A framework for overbooking and cancellations in airlines is put forth by Sulistio et al. [150] They introduce different methods, that are based on probability (a fixed overbooking percentage), risk aversiveness (overbooking as a function of risk) and service level (having the expected number of denied boarding passengers less than some level). Through simulations, effective overbooking techniques are shown to be able to generate up to 9% of additional revenues, which highlights the importance of accurate cancellation forecasts.

Part II

THE REAL PRACTICE OF REVENUE MANAGEMENT

3 EXPLOITATIVE CUSTOMER BEHAVIOR

Abstract

Customer behavior is rarely studied in the literature. Customers are often seen as "honest", and assumed they do not engage in exploitative behavior. In his Chapter, an overview of how pricing works in practice is given. Next, several exploits are introduced that are a result of the legacy of pricing of the 1980s. These exploits include avoiding fare restrictions, one-way pricing, and exploiting change and cancellation policies. Next, it is shown how customers can exploit these airline pricing strategies and provide discussion whether these exploits can be currently fixed (two of these exploits can be solved using the algorithms proposed in Chapter 8 and 9), may be stopped in the future, or may never be avoided.

This chapter is based on [17].

3.1 INTRODUCTION

Segmenting your customers properly is arguably the road to success in any industry. Having segmented your customers, creating a fitting, unique value proposition tailored to these groups will ultimately lead to maximum business performance. Different industries use different ways to segment their customers.

In the airline RM case, continued reliance on legacy systems means that the tools to segment customers are identical to those in the 1980s. These tools include requiring a customer to spend a minimum of days to spend at a destination before returning, needing to book a certain number of days in advance, and more. According to Botimer [97], working together with Delta Airlines, claims that these instruments "have been found effective by airlines in preventing passengers with a high values of willingness to pay from purchasing lower-priced fare products".

However, as technology involved, and us humans entered the information age, pricing engines have made minimal improvements since the inception of RM. The practical implications of these legacy ways of segmentation have not previously been discussed. The role of the internet has played a major role, as it has given the customer the opportunity, as well as made it much easier to exploit an airline's pricing structure.

As we found in Chapter 2, most of the work is done in the field of (dynamic) pricing and describes how to optimally set prices. There has not been any research in the exploits that the RM problem, in particular the subject of pricing, brings along. We define an *exploit* as a way for a customer to get access to a lower fare than the airline would like her to.

Many of the publication in the field of RM and pricing have focused on forecasting and optimisation. In the end, a combination of RM and pricing is required to extract optimal revenue. However, algorithms proposed in the literature often make assumptions that may be considered very crude. One example of this in most optimisation algorithms, is that it is assumed that there are no cancellations. Another typical assumption is that fares are fenced (segmented) optimally.

However, even when all assumptions are correct, the current way of RM and pricing provides a number of exploits that have not exposed in literature before. In this chapter, we provide eight such exploits, and discuss whether these can be combatted through either RM or pricing, or not at all.

Every exploit is explained and then illustrated through an example. All of the examples are set up using real data, accessed from publicly available global distribution systems (GDS).

This chapter is organized as follows. In Section 3.2, we provide an overview of how pricing works in practice. We discuss how airlines have traditionally, and up to this day, segment their customers. We have then split exploits in three categories. The first category, avoiding fare restrictions, is discussed in Section 3.3. Next, we discuss customer behavior, specifically cancellations and changes, in Section 3.4. Finally, we discuss exploits for one-way tickets in Section 3.5. We provide a discussion in Section 3.6 where we separate these exploits in those that can be fixed through RM and pricing, those that cannot be stopped as a result of legacy systems in place and those that cannot be stopped at all. We provide conclusions in Section 3.7.

We must stress that most of the abuse discussed in this chapter violates the condition of carriage of most airlines. Some airlines monitor and respond to this abuse more actively than others [16].

.

3.2 PRICING IN PRACTICE

Airlines use different ways to segment their customers. A product offered by an airline is a combination of minimum and maximum stay requirements, advance purchase requirement, seasonality, combinability, stopover possibility and more.

A *minimum stay requirement* is the number of days a customer is required to stay at their destination before returning. The length of stay is calculated by taking the difference between the date of the outbound and inbound journey. If this difference, expressed in number of days, exceeds the minimum stay requirement, this fare is available to the customer. Typically, minimum stay requirements is used to segment leisure from business customers. A customer staying at their destination for a single day before returning home is more likely to be a business traveller. With this in mind, cheap fares typically have a minimum stay requirement, so these are not available to business travellers, who often have higher budgets.

Maximum stay requirements are typically used to prevent customers booking the "wrong" direction. Similar to minimum stay requirements, they are calculated by taking the difference between the date of the outbound and inbound journey. It is checked whether this difference, again expressed in days, is smaller than some set value. Consider a ticket from London to Amsterdam. A Briton travelling to Amsterdam is happy to spend a few days in Amsterdam, while a Dutchman will most likely spend much longer in Amsterdam. Limiting the time one can spend at the destination ensures proper segmentation.

It is typically thought that customers purchasing tickets close to their departure dates have a higher willingness to pay: they "have" to travel. For this reason, airlines incorporate advance purchase requirements. Such a requirement is met if the difference between departure date and the current date exceeds a specified value.

Seasonality is perhaps the most intuitive way to segment customers. Seasonality rules ensure that fares are only available for certain departure dates. In practice, this means that seasonality rules restrict lower fares to be available during peak travel dates, such as school holidays or Christmas.

Combinability conditions specify what fare can be combined with what other fare. These situations arise when a fare differs for outbound and inbound portions of an itinerary.

Stopovers are offered by airlines to allow customers to spend time in between (connecting) flights, and therefore, to visit an additional city for a given fee. Some governments subsidize airlines to have them offer a free stopover to customers and use this as a way to boost tourism. Consider an example from London to Tokyo via Amsterdam. Rather than connecting directly, a customer may decide to book a stopover and spend some time in Amsterdam before continuing his or her journey to Tokyo. Airlines typically impose restrictions on the length of the stopover to avoid abuse. A customer staying in Amsterdam for six months before continuing to Tokyo is most likely not using the stopover as it was intended but rather engaging in abuse.

A *fare basis* encapsulates these conditions, as well as other conditions not mentioned here and typically starts with the letter of the associated fare class. A fare basis to fare class mapping is many-to-one: there can be more than one fare basis for the same fare class. One example is two different fare bases, with one allowing a weekend departure and one that doesn't, all else being equal. These fare bases then map to the same fare class. These fare bases are submitted to a Global Distribution System (GDS) which makes sure that these fares are available in different channels. The objective of RM is then to determine how many seats to make available for these different fare classes. A fare is only available for sale if all fare conditions are met and there is availability for that fare class.

In most airlines, fares are set by analysts. Involvement of mathematics ranges from advanced, providing recommendations how to set fares; to low, providing decision support; to none at all. Pricing analysts often argue that pricing is an "art", not a "science" that requires market knowledge to set fares and fare conditions appropriately. There remains to be a large amount of human involvement in setting fares.

Table 3.1 shows an example of the fare structure for economy class return fares from Tokyo to Dubai on Emirates Airline for travel dates in October 2020. The information in this table and tables to come are part of the public domain and retrieved from ExpertFlyer [151].

Fare Basis	Airline	Class	Fare	Min/Max Stay	Adv Pur
TLXJPJP1	EK	T	510	03 / 1M	7
LLXHPJP1	EK	L	548	03 / 1M	7
QLXHPJP1	EK	Q	610	03 / 1M	7
KLXHSJP1	EK	K	672	03 / 1M	3
ULXESJP1	EK	U	733	03 / 1M	3
BLXESJP1	EK	B	828	03 / 1M	3
MLXESJP1	EK	M	925	03 / 1M	0
WLXESJP1	EK	W	1022	03 / 1M	0
RLXEFJP1	EK	R	1407	03 / 4M	0
ELXEFJP1	EK	E	1695	– / 4M	0
YLXRFJP1	EK	Y	2080	– / 12M	21

Source: ExpertFlyer [151]

Table 3.1: Fares NRT - DXB, return

The lowest fare is priced at $510, and this requires a minimum three day stay in Dubai before returning in Tokyo. Customers are allowed to stay up to a month in Dubai before returning back to Tokyo for this fare to be available to them. Bookings need to be made at least seven days before their departure date. Note that there is no difference in fare conditions between *Q*, *L* and *T* class: the only difference is its fare value. These classes belong to what is commonly known as a *fare family* or *fare brand*, which we introduced in Chapter 2.

Once the departure date is less than seven days away, the lowest fare available for purchase is *K* class. The minimum and maximum stay requirements are similar to those of the *QLT* fare family, but can be purchased up to three days before departure (rather than seven). Once the departure date is closer than two days away, *M* class is the lowest available.

All the above assumed a customer is staying between three days and a month in Dubai. If a customer intends to stay less than three days, regardless of what time before departure a fare is purchased, *E* class is the lowest class available. Similarly, a customer intending to stay for more than four months is required to buy *Y* class, but this needs to be purchased at least three weeks before departure.

Notice that this fare structure means that the airline does not offer a fare for someone booking less than 21 days before departure, intending to stay more than four months in Dubai. After all, a customer cannot purchase *E* class because while the advance purchase requirement is met, the maximum stay requirement is exceeded. On the other hand, a customer cannot purchase *Y* class, because while the maximum stay requirement is met, the advance purchase requirement isn't. This example of human involvement mentioned above is most likely an error by the pricing analyst: a customer should always be able to buy an itinerary if purchasing the most expensive fare an airline offers.

3.3 AVOIDING FARE RESTRICTIONS

In this section, we will explore exploits that make it possible to avoid fare restrictions. We use the terminology of "exploit", since a customer is able to circumvent a restriction imposed by the airline. This can be achieved by abusing the stopover rule, by purchasing a ticket from a different POS than their own or nesting multiple bookings.

3.3.1 *Seasonality and stopover abuse*

Consider the fare structure from Kuwait to Amsterdam via Dubai, as shown in Table 3.5. Fares range from $314 in *T* class to $1482 in *Y* class. The *T* class product has the most restrictions: it needs to be purchased at least three days before departure and a customer can stay between three days and four months in Amsterdam. These restrictions are gradually relaxed up to *R* class - note that the *Y*, *E* and *R* classes form a fare family with the same conditions (none) but a different fare. There are seasonality restrictions for the fares shown in Table 3.5: these tickets can only be purchased for specific travel dates.

Seasonality is determined by the departure date of the first sector. Consider the fare rules for a fare basis, shown in Table 3.2. It states that this fare is available for departure dates between September 20, 2019 and June 15, 2020; and June 23, 2020 and July 20, 2020.

SEASONALITY	PERMITTED 20SEP19 THROUGH 15JUN20 OR 23JUN20 THROUGH 22JUL20 ON THE FIRST INTERNATIONAL SECTOR. SEASON IS BASED ON DATE OF ORIGIN.

Table 3.2: Seasonality of fare basis TXNVPKW1

The airline has put these fare rules because either traffic from Kuwait to Amsterdam is at its peak after July 22, or all traffic crossing from Dubai to Amsterdam may be high after this date. This is most likely because of the start of school holidays, occurring all around the world. Customers from different POS's all compete for a seat on the same Dubai to Amsterdam flight.

Should a customer still want to travel after July 22, this fare won't be available. In fact, the price jumps from $314 to $482 for departures after this date. However, there is a way to still purchase this fare: by purchasing an itinerary where the first sector occurs before the limits of the seasonality. An example is given in Table 3.3.

Date	Route
22-Jul	KWI - DXB
23-Jul	DXB - AMS
28-Jul	AMS - DXB
30-Jul	DXB - KWI

Table 3.3: Example itinerary

In Table 3.3, the first sector departs before the end of the seasonality (July 22). Because the fare is based on the date of the first sector, this combination of dates becomes eligible to purchase this fare, despite the DXB-AMS sector occurring outside the seasonality. In fact, this is possible without booking a stopover. As long as the departure date of the first sector is on or before July 22, this fare is available, regardless of the departure date of the connecting sector (DXB-AMS). In this way, a customer is able to get a low season fare for a high demand departure date.

3.3.2 *Point of Sale abuse*

Point of Sale (POS) abuse happens when a customer purchases a ticket originating from a different country than she is "supposed" to. Consider a customer based in Dubai. Since the customer is the United Arab Emirates, the POS is "supposed" to be AE. Traditionally, brick and mortar travel agents only used to have access to the POS of the country they operate in. For example, a travel agent in AE only had access to tickets offered by airlines for POS AE.

However, since the rising of the e-tickets, it is now very easy for this customer to purchase a ticket for a different POS. This could open the door for potential abuse: customers now have access to tickets they are traditionally not "supposed" to.

Consider this customer wishing to travel from Dubai to Amsterdam, spending a week there. The fares offered by the airline are shown in Table 3.6. This customer is only interested in obtaining the lowest fare, and does not mind change or cancellation policies. The intended departure date is further than three days away, and since both the minimum stay and maximum stay requirements are met, this customer is eligible for *K* class. This customer is "supposed" to buy this fare, since the customer is based in the United Arab Emirates.

To illustrate how this abuse works in practice, now review the fares for Kuwait to Amsterdam via Dubai. These are shown in Table 3.5. First, note that POS KW has fares available for *QLT* fare classes, while these are not available for POS AE for these departure dates. Second, note that up to *E* class, the booking classes for POS KW are roughly $200 lower like-for-like.

Nothing stops a customer from purchasing a KWI-AMS fare, as listed in Table 3.5. Suppose a customer purchases the *T* class fare, at a fare of $314. A sample itinerary is given in Table 3.3. However, as we will see in Section 3.4.1, segments should be flown in sequence and if a customer fails to show up for any segment any subsequent segments will be cancelled. This is one method airlines use to curb this abuse.

This means, in this case, that the customer has to fly the Kuwait to Dubai sector before flying the Dubai to Amsterdam sector. Therefore, the customer needs to purchase a ticket from Dubai to Kuwait first to complete the final itinerary. The fares for Dubai to Kuwait are listed in Table 3.7. By definition, these fares do not have a minimum or maximum stay (it is a one-way fare, after all). There are no advance purchase requirements either, which means that *T* class is available for purchase. Therefore, the total price is $314 + 52 = \$366$, which offers 52% of savings over booking a return DXB-AMS, POS AE flight, priced at $768. The final itinerary is given in Table 3.4.

Date	Route	Booking
01-Mar	DXB - KWI	1
01-Mar	KWI - DXB	2
02-Mar	DXB - AMS	2
13-Apr	AMS - DXB	2
14-Apr	DXB - KWI	2

Table 3.4: Example of final itinerary. Note that this contains two bookings: one booking DXB-KWI, POS AE and one booking KWI-AMS (via DXB), POS KW

Note that in the example of Table 3.4, the customer flies to Kuwait and back on the same day. However, since it is impossible to segment a one-way ticket, it is possible to spend one day (or more) in Kuwait before returning back to Dubai. We have chosen for these date combinations as this minimizes overall travel time.

Fare Basis	Airline	Class	Fare (USD)	Min/Max Stay	Adv Pur
TXNVPKW1	EK	T	314	03 / 4M	3
LXNVPKW1	EK	L	363	03 / 4M	3
QXNVPKW1	EK	Q	416	03 / 4M	3
KXNVSKW1	EK	K	482	03 / 4M	3
UEXESKW1	EK	U	667	– / 6M	0
BEXESKW1	EK	B	736	– / 6M	0
MEXESKW1	EK	M	809	– / 6M	0
WEXESKW1	EK	W	888	– / 6M	0
REXRFKW1	EK	R	1000	– / 12M	0
EEXRFKW1	EK	E	1185	– / 12M	0
YEXRFKW1	EK	Y	1482	– / 12M	0

Table 3.5: Fares for KWI-AMS (via DXB) return, POS KW

Fare Basis	Airline	Class	Fare (USD)	Min/Max Stay	Adv Pur
KLXESAE1	EK	K	768	03 / 4M	0
ULXESAE1	EK	U	860	03 / 12M	0
BLXESAE1	EK	B	956	03 / 12M	0
MLXESAE1	EK	M	1051	03 / 12M	0
WLXESAE1	EK	W	1146	– / 12M	0
RLXRFAE1	EK	R	1269	– / 12M	0
ELWRFAE1	EK	E	1397	– / 12M	0
YLXRFAE1	EK	Y	1588	– / 12M	0

Table 3.6: Fares for DXB-AMS return, POS AE

Fare Basis	Airline	Class	Fare	Min/Max Stay	Adv Pur
T1SOPAE1	EK	T	52	–	0
LLOOPAE1	EK	L	98	–	0
QLOOPAE1	EK	Q	128	–	0
KLOOSAE1	EK	K	166	–	0
USSOSAE1	EK	U	207	–	0
BSSOSAE1	EK	B	245	–	0
MSSOSAE1	EK	M	280	–	0
WSSOSAE1	EK	W	324	–	0
ROOWFAE1	EK	R	395	–	0
EOOWFAE1	EK	E	477	–	0
YOOWFAE1	EK	Y	585	–	0

Table 3.7: Fares for DXB-KWI one way, POS AE

3.3.3 *Nesting bookings*

Nesting of bookings is possible when one intends to travel at least twice with the same airline. The idea is to purchase two tickets of opposing POS's and flying these ticket nested. This way, one is able to circumvent minimum stay requirements. Refer again to Table 3.6, which contains the fares for Dubai - Amsterdam return. Table 3.8 show the fares for Amsterdam - Dubai return.

First, notice that fares offered for AMS-DXB (POS NL) start at class T, while fares offered for DXB-AMS (POS AE) start at class K. It should be noted that POS AE does offer class T, but these are not available for the travel dates chosen. There can be many reasons for this, and this is a good example of how pricing differs for different POS. Another observation is the difference in upsell amounts by class. The fares for B, U and K classes are similar between POS NL ($761) and POS AE ($768), but from W class upwards the differences become bigger. The most expensive class, Y, is almost double for POS NL compared to POS AE.

Consider the minimum stay requirements in Table 3.6 and 3.8. Note that for DXB-AMS, the first fare that is available in this customer's case is W class, priced at $1146. If a customer does not use nesting, the total amount for these two trips is $1146 * 2 = \$2292$. However, now consider booking a return DXB-AMS in K class, priced at $768. Next, this customer purchases a AMS-DXB return in T class at $378. Verify that all fare rules are met.

The resulting itinerary is shown in Table 3.9. Note that for the first trip, we use the first sector of Booking 1 (DXB-AMS return POS AE ticket) and fly back on the first sector of Booking 2 (AMS-DXB return POS NL ticket). Since these are two separate tickets, there is no way an airline can force a minimum stay requirement. Two weeks later, when this customer intends to travel to Amsterdam again, she uses the second sector of Booking 2. The minimum stay requirement for this Booking 1 is easily beaten. Lastly, the customer flies back to Dubai using the second sector or Booking 1. This example also shows the importance of the maximum stay requirement, introduced in 3.2. If there was no 1 month maximum stay requirement, we could use this itinerary to schedule a trip in March and, for example, in September.

The total charge for these nested bookings is $768 + 378 = \$1146$, a saving of exactly 50%.

Fare Basis	Airline	Class	Fare	Min/Max Stay	Adv Pur
TLEAPNL1	EK	T	378	05 / 1M	5
LLEAPNL1	EK	L	602	03 / 1M	5
QLEAPNL1	EK	Q	674	03 / 1M	5
KLEESNL1	EK	K	761	04 / 4M	3
ULEESNL1	EK	U	869	03 / 12M	
BLEESNL1	EK	B	1004	03 / 12M	
MLEESNL1	EK	M	1200	03 / 12M	
WLEESNL1	EK	W	1407	03 / 12M	
RLRZFNL1	EK	R	1652	– / 12M	
ELRZFNL1	EK	E	2324	– / 12M	
YLRZFNL1	EK	Y	2936	– / 12M	

Table 3.8: Fares for AMS-DXB return

Date	Route	Booking
01-Mar	DXB - AMS	1
02-Mar	AMS - DXB	2
13-Mar	DXB - AMS	2
14-Mar	AMS - DXB	1

Table 3.9: Example itinerary

3.3.4 *Minimum stay abuse*

Consider again the example in tables 3.5 and 3.6. Suppose a customer based in Dubai wants to travel for a one-day trip from Dubai to Amsterdam, departing December 20 and returning December 21. If we look at table 3.6, the lowest fare offered by POS AE that allows a one day stay is *W* class, priced at $1146.

The length of stay is calculated by taking the difference in dates between the departure date of the first sector and the departure date of the third sector, regardless of stopover presence. This detail may be subtle, but it is one that opens up the possibility to abuse.

To illustrate how a customer may benefit, consider the following itinerary, given in Table 3.10. This itinerary has two stopovers: one seven day stopover in Dubai, from December 13 to December 20; and one seven day stopover from December 21 to December 28. This is allowed as put forward in the fare rules, which are shown in Table 3.11. Note that this fare allows for two free stopovers with a maximum of seven days. The length of stay is calculated as being eight days: the difference in dates of the AMS-DXB and KWI-DXB sector, despite the customer spending only one day at their actual destination (Amsterdam). For this reason, this itinerary will be priced in *T* class at a fare of $314, which is shown in 3.5. Note that booking this itinerary, KWI-AMS will save the customer 72% compared to booking DXB-AMS.

Date	Route
13-Dec	KWI - DXB
20-Dec	DXB - AMS
21-Dec	AMS - DXB
28-Dec	DXB - KWI

Table 3.10: Example itinerary

STOPOVERS	3 STOPOVERS PERMITTED ON THE PRICING UNIT LIMITED TO 2 FREE AND 1 AT KWD 30.000. 2 FREE IN DXB 1 AT ANY POINT AT KWD 30.000. A STOPOVER MAY NOT EXCEED 7 DAYS.

Table 3.11: Fare rules of TXNVPKW1, POS KW

3.4 CUSTOMER BEHAVIOR

In this section, we describe cancellations, changes and rebooking by customers. The next sections are not exploits, but are opportunities for the customer to minimize their cost for the actions they wish to take. We discuss the risk the cancellation policy brings to the airline. Next, we discuss how a customer can benefit from fare movements by changing their booking at a specific time. This is followed by a strategy in which a customer rebooks to a lower fare class when flexibility is no longer necessary.

3.4.1 *Cancellations*

In this section, we demonstrate the danger of cancellations to the legacy (non-low cost airline) airline. To illustrate this, consider the penalty cause of a random fare basis. Note that this clause is consistent across every fare basis, and not specific to a given fare class. This is shown in Table 3.12.

PENALTIES	CHANGES ANY TIME CHARGE USD 100. ... A NO-SHOW FOR A FLIGHT IS CONSIDERED WHEN A PASSENGER FAILS TO REPORT AT THE AIRPORT AS BOOKED ONE HOUR BEFORE DEPARTURE OF THE SCHEDULED FLIGHT. ... FAILURE TO UTILISED TICKET AS BOOKED ON ANY SEGMENT OF THE ITINERARY WILL RESULT IN ALL SUBSEQUENT SEGMENTS OF THE ITINERARY BEING CANCELLED. IN SUCH CASES ONLY NO-SHOW FEE WILL APPLY AND NOT BOTH.

Table 3.12: Cancellation clause of fare rules

Table 3.12 includes two clauses that we would like to highlight. First, a customer is considered a no-show if she does not show up at the airport one hour prior to departure - this coincides with the closure of the check-in time of this airline. This, combined with the earlier clause that mentions that a change is possible any time, means a customer is able to avoid a no-show penalty as long as their itinerary is changed or cancelled at least one hour and one minute before departure. This puts the airline at significant risk.

Place yourself in the customer's perspective. *Why would I change or cancel my booking earlier than I am required to? Suppose it is now January 1 and my flight is scheduled on April 1 at 3PM. I already know that I may have to change my plans on April 1, as I have already received an invitation for a business meeting at 3PM. However, this is three months away and meeting times may change. Why would I not wait until the very last minute, say 12PM, the day itself, to move my booking?*

Customers are not given incentive to move bookings any earlier. Therefore, this exposes a lot of risk to the airline - if this customer does decide to move their booking three hours prior to the flight, the probability of selling another seat in that time frame is close to zero. This is, of course, the reason why airlines overbook their flights, to protect themselves against having empty seats. However, having time-dependent cancellation policies would make estimating overbooking levels easier and more accurate. We provide a framework for estimating these probabilities in Chapter 7 and provide an optimisation method in Chapter 9.

Second, it is important to stress the second part of the fare rules. These state that when a customer fails to use any segment of an itinerary, any subsequent segments will result in cancellation. In practice, this means that if a customer fails to fly or skips a segment, any remaining segments will be cancelled. For example, consider the example of Section 3.2. Here, a customer is looking to purchase a ticket Tokyo to Dubai return. Should a customer fail to show up for the Tokyo to Dubai sector of this ticket, the return portion of this ticket, Dubai to Tokyo will be cancelled automatically. This implies that all segments in a ticket should be flown in the sequence that they are booked. It is therefore impossible for a customer to book this itinerary and only fly the Dubai to Tokyo portion.

3.4.2 *Changing bookings*

Changes can be made to tickets where fare conditions allow the customer to do so. There is a distinction between a change *before* and *after* departure of the first segment. If a ticket is changed before the first segment is flown, this ticket is labelled as fully unutilised and a change to this fare is processed as if this customer purchases a new ticket (plus any change fee associated with this fare). On the other hand, after the first segment is flown, changes are processed using historical fares. This introduces an interesting concept for customers.

Consider a customer who purchases a ticket from Kuwait to Amsterdam, with a 14 day stopover in Dubai: the ticket consists of a coupon from Kuwait to Dubai on January 1, and connects from Dubai to Amsterdam on January 15. This ticket was purchased on July 1, for booking class T, allows for changes for $50 and cancellations at $100, at a fare of $250. It is now December 1, and the fare for booking class T has increased to $350. Availability has remained the same, and booking class T is still available.

A customer intends to change the Dubai to Amsterdam segment to January 14 while keeping the Kuwait to Dubai sector as is. It is in the customer's best interest to make this change after flying the Kuwait to Dubai sector. In this case, the ticket is partially used and this customer pays a change fee plus a fare difference based on the historical fare. Since T class is still available, the historical fare that is referenced is the old fare of $250. Therefore, the customer pays a change fee of $50 plus a fare difference of $0, amounting to a total of $50. Should the customer make the difference before flying the Kuwait to Dubai sector, the ticket is fully unutilised and she pays the same change fee of $50 plus a fare difference of $350 - 250 = \$100$, for a total of $150.

If a price has increased for a given fare basis has increased, it is therefore in the customer's best interest to fly the first sector if possible, then making any changes.

3.4.3 *Rebooking into lower fare classes*

Consider again the fare structure of Dubai to Sydney return fares, as shown in Table 3.13. The fare conditions for *L* class in this example indicate that a fare change costs $110, while this class cannot be cancelled. *K* class is the least expensive class that can be cancelled at $110, while changing this class costs $90. *R* class is the cheapest class that can be changed and cancelled without any penalties.

Fare Basis	Airline	Class	Fare	Min/Max Stay	Adv Pur
LXEEPAE1	EK	L	762	03 / 4M	5
QXEEPAE1	EK	Q	904	03 / 4M	0
KXEESAE1	EK	K	1318	03 / 6M	0
UXEESAE1	EK	U	1440	– / 6M	0
BXEESAE1	EK	B	1568	– / 6M	0
MXEESAE1	EK	M	1694	– / 6M	0
WXEESAE1	EK	W	1822	– / 6M	0
RXRZFAE1	EK	R	2010	– / 12M	0
EXRZFAE1	EK	E	2200	– / 12M	0
YXRZFAE1	EK	Y	2451	– / 12M	0

Table 3.13: Return fares, DXB SYD

Suppose it is now January 1, and this customer intends to travel to Sydney on May 1, returning on May 5. Assume that the RM system has made all these classes available for sale, so this customer is free to choose whatever class she intends.

In this case, this customer meets the conditions for every fare basis in Table 3.13. After all, this customer stays at their destination for four days, which exceeds the minimum stay and is less than the maximum stay for all fare bases. There are four months left until departure, so even the most restrictive advance purchase requirement (five days) is met. The customer is still not sure whether she can actually travel on May 1, so she intends to either purchase *K* class at $1318 or *R* at

$2010 class, as the flight is still four months away and she requires some flexibility.

This also illustrates why it is very unappealing for customers to buy R class. The fare difference amounts to $2010 - 1318 = \$692$. This means a customer is able to change a ticket six times and cancelling ($90 * 6 + 110 = \$650 < 692$) and would still be better off than purchasing a R class ticket that offers free change and cancellation. For this reason, this customer decides to purchase K class at $1318.

Suppose we are now one week before departure, and this customer is now sure that she is able to travel on these dates. Therefore, the flexibility that K class offers is no longer required. Nothing is stopping a customer to cancel K class, receive a refund for $1318 - 110 = \$1218$, and buying a new L class fare at $762. This enables the customer to save $456.

In Chapter 8 we provide a dynamic programming formulation that explicitly models customers that purchase the lowest available fare. This is later extended to customers that postpone their decision, possibly benefiting from availability that opens up close to departure.

3.5 ONE-WAY TICKETS

One-way fares are typically relatively expensive for legacy airlines. These fares are high, as it is typically thought customers that purchase these products are most likely relocating. The cost of relocating is typically burdened by the employer and therefore it is a classic case of "charge what you want". However, there are other reasons why one-way tickets are expensive: the airline loses the ability to segment the market. After all, a minimum and maximum stay do not exist for a one-way ticket. Advance purchase restrictions are still applicable of course, but a minimum stay requirement is often thought of as the most powerful tool that the airline has at their disposal. Therefore, the answer is to price these fares relatively high. In the following sections, we will explore how a customer can potentially circumvent these restrictions.

3.5.1 *One way abuse*

One way fares are often relatively expensive, and can be more expensive than a return fare in some instances. There can be many reasons for it. Traditionally, airlines charged a lot for these fares as customers that purchase one way fares are often thought of being relocating, and therefore, typically reimbursed by their company. Moreover, segmentation for one way is difficult: by definition, you cannot apply a minimum or maximum stay restriction.

In this example, we will look at fares between Dubai and Sydney. Table 3.13 contains the return fares, Table 3.14 show the one way fares. First observe that one-way fares are only offered from class *U* upwards, while return fares are offered from *L* upwards. Note that these fares offer no segmentation: with the exception of *W* class, there is no difference in advance purchase requirement.

Fare Basis	Airline	Class	Fare	Min/Max Stay	Adv Pur
USSOSAE1	EK	U	1187	–	0
BSSOSAE1	EK	B	1277	–	0
MSSOSAE1	EK	M	1367	–	0
WSSOSAE1	EK	W	1451	–	5
ROOWFAE1	EK	R	1588	–	0
EOOWFAE1	EK	E	1721	–	0
YOOWFAE1	EK	Y	1895	–	0

Table 3.14: One way fares, DXB-SYD

A customer who wishes to purchase a one-way to Sydney would be wise to purchase a return fare in *L* class, and simply not show up for the return sector. This particular fare, $LXEEPAE1$, is non-refundable, and thus, the customer will not be given a refund for the return sector. Refunds, in general, are dependent on whether the ticket is fully unutilised or not. A refund for a fully unutilised ticket is clear: the refund is equal to the fare value, minus a cancellation penalty. For a partly utilized ticket, after flying one or more sectors, refunds are processed as mentioned in Table 3.15 (refund calculations are independent of POS or OD).

REFUNDS	A2.AFTER DEPARTURE / PARTIALLY UTILISED TICKETS - AFTER COMMENCEMENT OF THE FIRST SECTOR OF THE JOURNEY. ... DEDUCT THE OW FARE OF EQUAL OR HIGHER AMOUNT THAN THE FARE PAID FOR THE PORTION OF THE JOURNEY PERFORMED IN THE SAME OR NEXT HIGHER RBD.

Table 3.15: Refunds for fare basis USSOSAE1, DXB-SYD

Consider a passenger buying a U class return fare, for \$1440, as shown in Table 3.13. The corresponding one-way fare is equal to \$1187, found in Table 3.14. After flying Dubai to Sydney, the ticket becomes partly utilized. The customer applies for a refund, as she has no intention to fly back. The purchased fare for the Dubai to Sydney sector is $\$1440/2 = \$720$. The one-way fare is now deducted, resulting in a refund of $\$720 - \$1187 = -\$467 \leq 0 = 0$. This example illustrates why booking a U class return fare, and applying for a refund is a less desirable strategy than booking a L class return fare.

However, there is potential for much greater abuse. This is done by exploiting combinability. *Combinability* is the concept of combining multiple fare bases to form (double) open jaws. An *open jaw* is a (return) ticket where the arrival airport of the outbound journey differs from the destination airport from the inbound journey. For example, a ticket Amsterdam to London, returning Birmingham to Amsterdam is considered an open jaw, since the return (Birmingham) departs from a different airport than the arriving airport (London). A *double open jaw* is where the departing airport of the outbound journey differs from the arrival airport of the inbound journey. For example, a ticket Amsterdam to London, Birmingham to Brussels is considered a double open jaw.

Open jaws are traditionally present to offer customers the flexibility of visiting two cities on a single ticket. While transportation has to be arranged between London to Birmingham in the example above, it is much cheaper than purchasing two separate tickets Amsterdam - London return and Amsterdam - Birmingham return. From a pricing perspective, the fare paid by the customer is a combination of half of the return fare Amsterdam - London, and half of the return fare Amsterdam - Birmingham. Suppose, for example, that the fare available for Amsterdam - London is \$100 and Amsterdam - Birmingham is \$80. Then, the fare for this open jaw is equal to $\$0.5 * 100 + 0.5 * 80 = \$90$.

While this is great for customers, it does, also, open up possibilities for abuse when one is interested in a one-way fare. Consider again the Dubai to Sydney case, and let's consider the $LXEEPAE1$ fare basis. The combinability of this fare basis is shown in Table 3.16.

COMBINATIONS	OPEN JAWS/ROUND TRIPS/CIRCLE TRIPS FARES MAY BE COMBINED ON A HALF ROUND TRIP BASIS -TO FORM SINGLE OR DOUBLE OPEN JAWS/ROUND TRIPS -TO FORM CIRCLE TRIPS A MAXIMUM OF TWO INTERNATIONAL FARE COMPONENTS PERMITTED. PROVIDED - COMBINATIONS ARE WITH ANY Y-FAE1/E-FAE1/R-FAE1/ W-SAE1/M-SAE1/B-SAE1/U-SAE1/K-SAE1/Q-PAE1/L-PAE1/ T-PAE1/X-RAE1/V-RAE1 TYPE FARES FOR CARRIER EK WITH ANY RULE IN ANY PUBLIC TARIFF.

Table 3.16: Combinability of fare basis LXEEPAE1

The fare rules indicate that this fare may be combined, to form a single open jaw, with any fare basis that ends in FAE1, FAE1, R-FAE1, W-SAE1, M-SAE1, B-SAE1, U-SAE1, K-SAE1, Q-PAE1, L-PAE1, T-PAE1, X-RAE1 or V-RAE1. The fare rules do not impose any restriction on the origin and destination of this fare. Now consider the fares for Dubai to Delhi, shown in Table 3.17.

Fare Basis	Airline	Class	Fare	Min/Max Stay	Adv Pur
XLXERAE1	EK	X	82	– / 3M	0
TLXEPAE1	EK	T	112	– / 3M	0
LLXEPAE1	EK	L	139	– / 3M	0
HASLRAE1	EK	H	152	03 / 1M	5
QLXEPAE1	EK	Q	169	– / 3M	0

Table 3.17: Return fares, DXB-DEL

The lowest fare basis is $XLXERAE1$, which is a fare basis that ends in RAE1. Therefore, this X class fare is combinable with the fare Dubai - Sydney with fare basis $LXEEPAE1$. This itinerary is shown in Table 3.18.

Date	Route	Fare Basis	Round Trip Fare
1 Mar	DXB - SYD	LXEEPAE1	762
13 Mar	DEL - DXB	XLXERAE1	82

Table 3.18: Example Itinerary: DXB - SYD (open jaw) DEL - DXB

Exploiting this, the fare offered to the customer is $0.5 * 762 + 0.5 * 82 = \$422$. This option offers substantial savings for the customer compared to purchasing a simple Dubai to Sydney return ($762) and even greater savings compared to booking a Dubai to Sydney one way fare ($1187).

3.6 discussion

In this chapter, we have introduced different possibilities for abuse in RM. We have grouped them into three categories:

1. Abuse as a result of legacy systems that cannot be stopped;
2. Abuse that can be stopped through fare rules;
3. Abuse that cannot currently be stopped.

Many of these issues arise from the reliance on old, legal systems. These systems are built in such a way that prevents airlines from fixing this abuse. Some abuse cannot be stopped in the pricing engines, but can be fixed by incorporating fare rules. Lastly, there is abuse that cannot be stopped using current systems.

First, we would like to highlight that the number of customers that engage in RM abuse is likely to be very low. One needs to have extensive knowledge of RM, and in particular, pricing to find these loopholes. We therefore find that our work is aimed at creating awareness to airlines where possible exploits could arise from, rather than investing in resources to find abusive customers.

Section 3.3.1 shows how one can use a stopover to start a journey in a low season, and continue a journey into a high season while still getting access to a low season fare. This illustrates a limitation of legacy systems.

Point of sale abuse, introduced in Section 3.3.2, shows how customers can benefit from fares from a different point-of-sale than their own. This technique relies on a one-way fare to start the journey, so this exploit may only be worthwhile is within "decent" distance of one's home airport.

It should be noted, however, that some of this abuse can be solved by using specific fare rules. For example, consider the one-way abuse example of Section 3.5.1. In this example, a customer is looking to fly a one-way journey from Dubai to Sydney but abuses the system by purchasing an open jaw Dubai - Sydney / Delhi - Dubai ticket. Sydney and Delhi are 6476 miles apart, so one can argue that offering the possibility for an open jaw for this itinerary shouldn't be allowed, as the number of legitimate passengers for an itinerary like this is extremely low and the possibility for abuse far outweighs the potential revenue of legitimate customers. Fare rules can prohibit combinability by area. The airline may consider that open jaws are only allowed within a specific region; for example, restricting open jaws to return fares from within the Asia Pacific region.

Some abuse cannot be stopped with current systems used by the vast majority of airlines. One would question why airlines simply do not move on to a more advanced system. In our view, this is a result of fear of change ("Why fix something that has worked for years?"), but, more importantly, because partnerships between airlines mean that systems need to be able to talk to each other. Many airlines have partnerships ranging from joint-ventures to code-sharing (selling seats on a different airline using your own flight number) to interlining luggage. Upgrading systems would likely result in compatibility issues at best, and a complete loss of communication.

Lastly, it should be noted that the majority of these exploits require a customer flying at least two extra sectors. Not every customer wants to spend his or her time to fly back-and-forth to an airport before flying to their actual destination, simply to obtain a lower fare.

Section	Objective	Prerequisite	Exploit used	Fix possible?
3.4.1	Prelonging the time before cancellations	None	Cancellation fee does not depend on time	Currently
3.4.2	Changing departure date for second or later flight	Customer needs to fly first sector	Historical fares that are typically lower	No
3.3.1	Travel to final destination in high season	Customer needs to fly first sector outside of high season	Seasonality is based on first sector	Future
3.3.2	Obtain lower fares for a return flight	Customer needs to book a one-way flight and fly these sectors	Fares for non-hub itineraries is typically lower	No
3.3.3	Avoid minimum stay requirements	Customer needs to be willing to travel twice	No restrictions possible between different bookings	No
3.3.4	Avoid minimum stay requirements	Customer needs to book a one-way flight and fly these sectors	Length of stay calculated based on date of first sector	Currently
3.5.1	Obtain lower fares for a one-way flight	None	Purchasing a return fare, using open jaw	Currently
3.4.3	Obtain lower fares once plans have materialized	Initial fare needs to be refundable	Easing of availability over time	Currently

Table 3.19: Overview of different exploits

3.7 CONCLUSION

In summary, we have the following observations and conclusions:

1. Legacy pricing systems and the rise of the Internet have made it possible for customers to purchase products that were previously unavailable.
2. Traditional segmentation methods, such as advance purchase, minimum and maximum stay requirements may be bypassed through different strategies.
3. Using these exploits have two benefits to the customer: significant savings, and he or she gains access to products that are not supposed to be available.
4. These strategies, or exploits, can be categorized between those that cannot be currently stopped, may be put an end to the in the future with better systems, and those that may never be barred.
5. Some of these methods breach the Condition of Carriage, with some airlines taking legal action to customers illustrating the financial damage airlines suffer.

4

SYSTEM AND CULTURE LIMITATIONS IN PRACTICE

Abstract

The theory of RM is often not applicable in practice. This is usually a result of RM system limitations and the culture in the department. In this chapter, we first provide a brief overview of the most popular optimisation method. Having identified this method, we note the inputs that are required. In the following sections, we show difficulties in the demand forecast inputs, pricing inputs, the actual optimisation and the distribution of the fare offering. In terms of demand inputs, we identify challenges in unconstraining, usage of DCPs, groups, cancellations, connections spanning more than a day, and determining what OD-pair a customer has purchased. Difficulties for pricing inputs include differences in fare value across distribution channel and how fares are distributed across different legs. For optimisation, we discuss infrequent optimisation, discretizing time, code sharing, one-way RM, load factor culture, segment limits, groups and denied boarding cost.

This chapter is based on [18].

4.1 INTRODUCTION

In this chapter, we will outline the limitations of RM systems and the RM culture in practice. The end goal of RM is to find an optimal policy that depends on what to achieve (maximize revenue, profit, forecast accuracy, and so forth). To start, we would like to briefly repeat the most popular optimisation technique that maximizes revenue.

We begin by repeating the traditional dynamic programming formulation, introduced by Talluri and Van Ryzin [65]. Consider:

λ_j is the arrival rate of class j, $j = 1, \ldots, J$;
f_j the fare of product/class j, $f_1 \geq f_2 \geq \ldots \geq f_J$;
x the remaining capacity;
t the time unit, $t = 1, \ldots, T$.

As before, demand follows Poisson process and suppose we have discretized time in such a way that in every time slice, we can have at most one arrival. Discretizing is done in such a way that the probability of having more than two arrivals is smaller than some chosen value ϵ (typically, $\epsilon = 0.05$). Note that $P(R(t) = p_j) = \lambda_j(t)$. When presented with an arrival, we need to decide whether to receive the current revenue, given by the random variable R_t, and move to the next time unit with one unit of capacity less, or reject this arrival request but have the same number of capacity in the next time unit. Therefore, introduce an indicator variable $u \in (0, 1)$, which is what we want to maximize over:

$$R(t)u + V_{t+1}(x - 1)$$

Now define $V_t(x_i)$ as the value function that represents the revenue-to-go given t units of time left and x units of capacity for a flight i.

$$V_t(x_i) = E\left(\max_{u \in (0,1)} R(t)u + V_{t+1}(x_i - u) \right) \qquad \textbf{(DPID)} \tag{4.1}$$

Note that this problem is formulated for a single flight: x is simply the capacity of one particular flight. It is then assumed that the network revenue is the sum of all individual flights:

$$V_t(\mathbf{x}) \approx \sum_{i=1}^{N} V_t^i(x_i) \tag{4.2}$$

in this case, x is now a N-dimensional vector with element i containing the remaining capacity of flight i.

Therefore, to solve the RM problem, we need:

1. To satisfy the conditions of the Poisson process, discretize time t;
2. Since no cancellations are assumed, estimate unconstrained demand free of cancellations, λ_j;
3. Obtain an estimate for a fare for every class j, f_j;
4. Solve the optimisation laid out in Equation 4.1;
5. Ensure fares are made available to the customer.

In practice, this means we need to answer the following questions:

- What is our (unconstrained) demand forecast?
- What are our fares?
- How do we determine the optimal policy?
- How do we ensure these fares are available to the customer?

In the next sections, we describe practical problems that will show that these steps are not as easy and straightforward as they may seem. We have separated this chapter in four sections, each answering one of the questions raised above: we will cover the challenges with demand forecasting in Section 4.2. Next, the difficulties with fares are being discussed in Section 4.3. This is followed by the practical problems of optimisation, in Section 4.4. The practice of distribution is discussed in Section 4.5. We will illustrate that these limitations can be grouped into four different categories: data storage and computation time, approximations with merit, approximations without merit, and culture. We provide a discussion in Section 4.6 and close with conclusions in Section 4.7.

4.2 DEMAND FORECASTING

In this section, we will look into answering the first question: **what is the unconstrained demand forecast?** This will be approached from different angles, each exposing how the theory of RM greatly simplifies the practice. First, we discuss difficulties in the process of unconstraining in Section 4.2.1. Second, we expose the practice of using the same data collection points (DCPs) across the networks for forecasting and optimisation in Section 4.2.2. Third, we discuss the difficulty of forecasting group booking demand in Section 4.2.3. Fourth, predicting cancellations and its limitations is discussed in Section 4.2.4. Fifth, in Section 4.2.5 we discuss the concept of a network departure date and connections and provide an overview of challenges this poses to the system. Sixth, we discuss how an OD-pair is derived from a booking in Section 4.2.6 and note how the system falls short.

4.2.1 *Unconstraining: class closure flag*

Before demand is forecasted, historical observations are unconstrained. This process, estimating demand in absence of any (optimisation) strategy, is conducted to predict the total (population) demand. Only observations for a given product (class) that are closed for sale are unconstrained. After all, if a class is available for sale, the true demand is recorded, and therefore, uncensored by definition. In practice, unconstraining is done by DCP. This is done because forecasting is also done by DCP, so it is natural to uncensor demand at this same level of aggregation.

Day	10	9	8	7	6	5	4	3	2	1	0
Available	1	1	1	1	1	1	1	1	1	1	0

Table 4.1: Booking class availability, 1 represents this class is available for sale, 0 indicates the class is closed for sale.

However, since a DCP (typically) consists of two or more days, this poses the question when a DCP is closed. Consider the example given in Table 4.1. Here, a DCP consists of the last 10 days before departure. A particular class is available for sale for 9 out of the 10 days, but is not available for sale on the last day of the DCP. The question now is: do we consider this DCP censored? In practice, there are multiple heuristics. One possible heuristic uses the majority of observations to determine whether the entire DCP Is open. The most crude approximation we have seen in practice only considers the status at the *last* day of the DCP, disregarding all other points.

4.2.2 *Network-wide DCPs*

The benefit of DCPs are two-fold. First, the variation in demand is reduced, as we aggregate a range days before departure. Second, it reduces the number of forecasts the system needs. In practice, most RM systems use between 10 and 20 DCPs. Assuming that bookings can me made 340 days before departure, this means that the number of forecasts is reduced by 34- or 17-fold, respectively. This, of course, greatly speeds up the computation time and makes frequent forecast updates possible. The importance of this was discussed in Chapter 2.

Within the DCP, it is assumed demand arrives linear in time. For example, if a DCP consisting of three days is forecasted with demand three, it is assumed that there is one unit of demand per day. In most RM systems in practice, the same definitions of DCPs are used across ODs, without a real need for it.

Consider a product for which we obtain a demand estimate of five in a DCP of length 5 days. With $\epsilon = 0.05$, this requires 98 time units. We can convert these time units to days by means of a linear spline. Let d be the number of days before departure which represents the start of the DCP. Then this results in values of $d{+}0, 0.041, 0.0812, 0.1224, 0.1633, 0.2040, 0.2445, \ldots$ Another product is expected to only receive a single unit of demand, and this results in the need for 20 time units. Converting this to days, we obtain $d{+}0, 0.2, 0.4, \ldots$. One of these vectors can then be used as a dictionary to look up a value closest to it. This illustrates the fact that different DCPs can be used by first converting time units to days before departure, finding the closest match, then using the index in the vector associated with this closest match. In the example above, the first element of the first vector is matched with the first element in the second vector (trivial). Next, the fifth element of the first vector is matched against the second element in the vector, and so forth. These can then be used in the optimization.

Note that in practice, this does not happen: a fixed number of time units is chosen. If a too small number of time units is chosen, the

Poisson process of having more than a single arrival per time slice is violated. If a too large number of time units is chosen, we unnecessarily create more forecasts than necessary. The main benefit of fixing time units is that there is no such thing as matching vectors - by definition the first element corresponds to every other first element.

4.2.3 *Group bookings*

RM systems typically exclude groups in its optimisation. After all, in Equation 4.1 we assume that for every time unit, there may be at most a single booking request. Groups are often dealt with through a different system, or a module outside the RM system. While for individual bookings automated control mechanisms like booking limits or bid price are common place, groups are often dealt with by an analyst in practice. Upon receiving a request, they gauge whether to accept the request based on the date, route and fare.

Group bookings differ from individual bookings in different ways. In practice, a distinction is made between ad-hoc and series requests. An *ad-hoc request*, as it implies, is a one-time request from a travel agent for a route or cabin. Series requests, on the other hand, are requests by travel agent that are repeated cyclically. There is also no notion of multiple fare classes for groups: they typically use a single booking class. For individual bookings, fares are fairly consistent over time within a booking class. This makes deriving a fare relatively easy. However, group fares can vary widely, yet are attributed to the same booking class. Consider for example a request for a flight that is only selling the third most expensive booking class. The airline may very well accept this request, despite the fact that the group fare typically being offered, is one that is at a discount of the equivalent available (individual) fare. Now consider another request, for a flight that is empty and the cheapest booking class is available. The airline may not accept this request at an entire different price point, yet demand is assigned to the same booking class. This causes significant difficulty in determining what the "group" fare should be in optimisation, and is one of the reasons why this is often left out of the optimization but done separately. Another problem is the high variability in group requests and therefore difficulty in forecasting group demand. This is why group bookings are sometimes seen as incremental revenue, and nothing more.

Group bookings pose significant risk to the airline. In practice, travel agents that coordinate group bookings exploit this by continuously negotiating with different airlines. More expensive airlines are dropped, and this often happens very late in the booking curve.

4.2.4 *Estimating cancellations*

In one of the most used optimization techniques in practice, which follows Equation (4.1), cancellations are assumed to be absent. Indeed, given a state x, the remaining capacity, we may move to the next stage with the same state x, or sell a unit of capacity and end up in state $x - 1$. Jumps to state $x + 1$, which would happen if cancellations are allowed, after all, a unit of capacity is returned, are not possible. This, in turn, means that the demand estimate for product j, λ_j, from (4.1) should be a forecast free of cancellations. In practice, this is often referred to a *net demand*. Calculating net demand is not as straight forward as it may seem. In practice, there are two different ways that are used: in the first method, demand and cancellations are forecast separately using historical observation. In the second method, historical net demand observations are calculated before these are used in the forecasting process. Since we assume λ_j follows a Poisson process, we must have $\lambda_j > 0$.

Consider the bookings and cancellations for a product for a certain departure date, which are shown in Table 4.2.

Type	DCP1	DCP2	DCP3	DCP4	DCP5
Bookings	5	2	0	2	1
Cancellations	2	1	3	1	1

Table 4.2: Example of bookings and cancellations for a given departure date. DCP 1 is the start of the booking curve.

Method 1: forecasting bookings and cancellations separately Suppose, for simplicity sake, that there is no variance in bookings and cancellations across multiple departure dates, so every departure date the same number of bookings and cancellations is recorded as shown in Table 4.2. Since there is no variance, the forecast is simply the values shown in this Table. Consider DCP3. Here, two bookings are forecast. Cancellations are forecast in isolation, and are predicted to amount to four for this DCP. The net demand, given by the difference between bookings and cancellations, is now -3. However, since we stated that we need to have $\lambda_j > 0$, this clearly cannot be an input into our optimization. This is the inherent danger of forecasting bookings and cancellations separately: by definition, a cancellation cannot occur before booking. Similarly, the number of cancellations cannot exceed the number of bookings.

In practice, this is solved by forecasting the number of bookings, and applying a cancellation rate. The cancellation rate is the percentage of the number of bookings that is expected to cancel. After all, a (non-negative) cancellation rate means the net demand, the result of the multiplication of bookings and the cancellation rate, cannot be negative, and we always find a positive value of λ_j.

However, this raises the next question: how is the cancellation rate calculated? The cancellation rate for DCP3 is undefined, for the same reason as raised above. Therefore, cancellation rates are typically based on cumulative demand. The cumulative number of bookings is given by $(5, 7, 7, 9, 10)$ while the cumulative number of cancellations is equal to $(2, 3, 6, 7, 8)$.

In practice, once a DCP completes, a cancellation rate is generated, which is a function of the cumulative bookings and cancellations up to that time. After DCP1 completes, the cancellation rate is given by the ratio of bookings and cancellations up to DCP1, which is $2/5$. The cancellation rate for DCP2 is calculated after this DCP completes, and is given by $(2+1)/(5+2) = 3/7$. This cancellation rate is applied to both DCP1 and DCP2. This cancellation rate is replaced for DCP3 by $(2+1+3)/(5+2+0) = 6/7$ and is assigned to DCP3 and previous DCPs. This processes is repeated up to DCP5, and the final cancellation rate, that is assigned to all DCPs, is given by $(2+1+3+1+1)/(5+2+0+2+1) = 8/10$. This, in turn means that the final net demand estimates are given by $5*(1-8/10) = 1$, $2*(1-8/10) = 0.4$, $0*(1-8/10) = 0$, $5*(2-8/10) = 0.4$, $1*(1-8/10) = 0.2$.

Method 2: forecasting using net demand

Next, we discuss the second method. Note that from Table 4.2, only for DCP1 we can calculate true demand. After all, because it is the first DCP, the cancellations in this DCP are all the result of bookings that were made in that same DCP. Now consider DCP2. From Table 4.2, it is impossible to know whether these cancellations correspond to bookings made in DCP1 or DCP2. Therefore, for this method more data is used, which includes the time of booking and time of cancellation for each booking. This is shown in Table 4.3.

Booking	DCP Booked	DCP Cancelled	Included
1	1	1	N
2	1	1	N
3	1	2	N
4	1	3	N
5	1	3	N
6	2	3	N
7	2	-	Y
8	4	1	N
9	4	-	Y
10	5	5	N

Table 4.3: Example of booking times and cancellation times for a given departure date. DCP 1 is the start of the booking curve. The included column indicates whether this data point is used for the net demand calculation for Method 2.

In this method, only bookings that are never cancelled are included in the calculation of net demand. The column in this Table indicates this. For example, all bookings that were made in DCP1 were cancelled: out of the five bookings, the first two were cancelled in DCP1, the third was cancelled in DCP2, and the fourth and fifth were cancelled in DCP3. Since five bookings were received in this DCP, but all of those five were cancelled in the end, the net demand for DCP1 is 0. In DCP2, two bookings were received, of which one is cancelled. In DCP4, this happens again so therefore the demand for these DCPs is 1 each. Lastly, the one booking that was made in DCP5 is cancelled, so the DCP5 net demand is equal to 0.

In practice, forecasting using this method works as follows. Once a DCP is completed, net demand is calculated as above. Next, this value is uncensored, should the class have been closed for sale. This is done as described in Section 4.2.1. Next, this value is used in a forecasting method.

4.2.5 *Connections and demand forecasting*

The RM problem is solved for one departure date individually, which we define as a network departure date. 23:59 is used as a cut-off for network departure dates, so this implies the network revenue for January 1 is optimized independently of January 2. In practice, since most approximate dynamic programming methods assume that the network revenue is the sum of individual flight revenue, effectively flights are being optimized. This, in turn, means that a flight departing at 3PM is optimized independently of a flight departing at 1AM, even though itineraries may cover both of these flights. While this may seem like a reasonable assumption at first glance, this causes a lot of problems for airlines that have next-day connections. Curfews in Europe and the United States mean that this is not an issue in most cases in those territories, but airlines in Asia and the Middle East suffer from this assumption. Consider the following itinerary:

Weekday	Date	Route	Flight Number	Dep. Time	Arr. Time
Sunday	22-Jul	AMS - DXB	EK148	15:20	23:59
Monday	23-Jul	DXB - SYD	EK412	01:40	22:25

Table 4.4: Example of itinerary spanning two network departure dates

What network date is assigned to this itinerary? Clearly, assigning both sectors of this PNR to a 22 July network departure date is wrong, since the second leg departs on 23 July. However, if this PNR is split into two network departure dates, we effectively split the (single) OD, AMS-SYD, into two different ODs, AMS-DXB on July 22 and DXB-SYD on July 23.

When the former strategy is used, in a way a more accurate view of demand is given: both flights are assigned one unit of demand within the same optimization. This is particularly problematic when one day is a weekday, while the other is a weekend day. In the example of Table 4.4, it is clear that the demand for EK412 is the result of passengers leaving from Amsterdam on a weekend day.

When the latter strategy is used, we assign demand correctly (that is, this booking counts as one for EK412 on 23 July, not on 22 July) but we lose the aforementioned OD effect.

Another issue arises from assigning fares. The leg fare for a given OD is weighted based on distance. For example, assume that the weekend fare for the itinerary in Table 4.4 is $1069, while the weekday fare is $962.10. A fare is defined by the date of the first outbound sector. For example, the itinerary in Table 4.4 is priced for a weekend fare, since the departure for the first leg is on a Sunday. Itineraries departing on a weekend typically incur an extra fee, and is a way an airline segments demand.

The distance of AMS-DXB is 3215 miles and DXB-SYD is 7481 miles for a total of 10690 miles. Then $321.50 of its fare is assigned to AMS-DXB and $748.1 to DXB-SYD. However, if we approach it using the second strategy, the week-day fare will be looked up and split, resulting in $673 for the DXB-SYD leg while AMS-DXB remains the same $321.50. This will result in errors in revenue accounting, since the sum of the leg-prorated fares no longer tallies with the (actually paid) OD fare of $1069.

One heuristic that is used in practice is to scale these fares so they do sum, but this is done linearly: in the latter case, we have a scaler of $1069/(321.50 + 673) = 1.075$, which results in a final leg fares of \$345.62 and \$732.38. Note that this causes the AMS-DXB proportion to be higher than previously calculated and the DXB-SYD lower.

4.2.6 *Determining OD-pair*

All PNRs are processed by the RM system before being used as historical observations in the forecasting module. One of the objectives is to determine what OD-pair was actually travelled by the customer. This may not be as straight forward, and involves manual (subjective) inputs. Consider the following itinerary:

Date	Route
1-Jul	KWI - DXB
2-Jul	DXB - AMS
28 -Jul	AMS - DXB
29 -Jul	DXB - KWI

Table 4.5: OD-builder, Example 1

Before this booking can be used in the forecaster, the system must define its OD. In this case, this PNR consists of two ODs that are split: KWI-AMS and AMS-KWI. Since this itinerary does not contain any stopovers, defining the OD is a straightforward process. However, now consider the following itinerary in Table 4.6.

Date	Route
1 -Jul	KWI - DXB
20 - Jul	DXB - AMS
21 - Jul	AMS - DXB
01 - Dec	DXB - KWI

Table 4.6: OD-builder, Example 2

One may argue that this is still a passenger travelling from Kuwait to Dubai on its outbound journey, with a long stopover in Dubai. However, at what point does a stopover constitute two separate trips? For the return sector, the stopover is over four months in length, so it may be assumed that this isn't a "true" stopover, but more likely to be a side trip. In this case, the OD is most likely to be defined as KWI-AMS, outbound; AMS-DXB and DXB-KWI inbound. The same concerns on network date addressed above apply. In practice, airlines lack a scientific approach to defining this cut off time and is defined by analysts. These cut off ranges from a week to a month. Next, consider the itinerary in Table 4.7.

Date	Route
1 -Jul	AMS - DXB
4 - Jul	DXB - BKK
10 - Aug	BKK - HKG
12 - Sep	HKG - DXB
12 - Sep	DXB - AMS

Table 4.7: OD-builder, Example 3

The customer spends three days in Dubai, one month and six days in Bangkok, and one month and two days in Hong Kong. How should this OD be split? In practice, most RM systems look at the time spent at a destination and use this this as the breakpoint to split the ticket. In this case, the since the customer spends the most time in Bangok, it will first be split here. Since the DXB-BKK leg occurs three days after the first leg, it is not split further. Therefore, the first OD from this PNR is AMS-BKK. Next, consider the last three legs. Since the customer spends more than a month in Hong Kong, the PNR is further split. The fourth and fifth sectors occur on the same day, so this is counted as a connection and no split is necessary. Therefore, this PNR is split into AMS-BKK, BKK-HKG and HKG-AMS.

4.3 PRICING

In this section, we will review the question: **what are our fares?** This may seem like a straightforward question. And, the theory of RM assumes that fares are given. However, in practice there are many intricacies that make this seemingly easy question very hard to answer. In Section 4.3.1 we go into detail and discuss our notion of "net fares". This is followed by Section 4.3.2, in which we discuss what are the actual fares that are used in the optimisation.

4.3.1 *Net fares*

In practice, there is no such thing as the "fare" for a particular product. This is because the fare received by the airline, after charges, fees, kickbacks, discounts is very different than fares actually filed or advertised. Moreover, fares change by type of customer: there are tens of different fares by the airline, which are offered to certain individuals: seaman, military, government, student discount, and many more. Each of these fares have some sort of discount (or surcharge) associated with them, but all these fares are aggregated when the fare for a product is calculated. We define the *net fare* as the fare that is actually received by the airline.

Airlines work with GDSs that distribute their fare offerings. In return, the airline pays a fee to the GDS provider. This is often done by segment. Should a ticket be sold via its own channel, there is no such fee. In this case, the net revenue for tickets sold through the distribution channel is equal to its ticket value. This is, however, the only exception when the net fare is equal to the fare quoted. If a ticket is sold via GDS, the fee it pays to the GDS is subtracted to obtain the net fare. Airlines often work with corporate contracts, which may offer kickbacks. It can, therefore, be argued that the net fare for corporate contracts should be (proportionally) adjusted for this kickback.

In practice, the airline now needs to decide what fares to include and exclude from its fare calculation. Ideally, the RM system should be made aware of these net fares. In practice, the gross fare value is used. We define *net revenue management (NRM)* the practice of RM, allowing for changes in fare between distribution channel. This could, for example, mean that the optimal strategy for NRM is to reject requests through the GDS for its last for sale, while it accepts the same request made through their own, direct channel as its new fare value is higher. In practice, however, contractual agreements between airlines and GDS, so-called *full content agreements (FCAs)* force the airline to offer the same (filed) fare through the GDS. This means that while the gross fare is the same, its net fare is lower as a result of the GDS fee.

4.3.2 *Fares for optimisation*

In Section 4.2.5, we discussed how demand is split when a connection occurs the next day. Similar to demand being split, the same is to be done for fares. In that section, we introduced an itinerary that consists of two flights: EK148 and EK412, let us assume that these are flights $i = 1$ and $i = 2$, respectively. This itinerary was offered for \$1069. Since in practice most airlines use the approximation introduced in Equation (4.2), the objective is to solve the flight-specific value functions from Equation (4.1). The challenge is identifying what fare to assign for the value function of flight $i = 1$ and $i = 2$.

It is clear that we may not assign \$1069 to both value functions. After all, Equation (4.2) no longer holds: this value function is then over-estimated by a factor two, since this fare is counted twice. Therefore, another way needs to be found.

In practice, the revenue accounting department is tasked to calculate the financial performance of every flight. When they are faced with this problem, the industry practice is to split these fares based on distance. The benefit of this approach is that the approximation holds: there is no double counting.

While this approach may seem reasonable for revenue accounting purposes, the downside for optimisation is undervaluing a relatively short leg and overvaluing a relatively long leg. Consider, for example, the itinerary LHR-AMS-SIN for $450. The distance of LHR-AMS is much smaller than AMS-SIN, resulting in only $30 assigned to LHR-AMS while $420 is assigned to AMS-SIN. For the same reason, splitting a fare evenly across flights satisfies the approximation but intuitively it does not seem "fair". Remember that different ODs are "competing" in every flight's value function, so a $225 fare component of a LHR-SIN may compete with a $500 AMS-SIN OD fare, which fare component for the AMS-SIN leg is equal to its OD fare.

Therefore, in practice the most common approach involves solving a linear program. In this linear program, the decision variable is how much to sell of a product j, subject to constraints that say the sum of the assigned decision variables must not exceed the capacity of every flight. The shadow prices (Lagrangian multipliers) are obtained. Recall that a shadow price is only positive if that constraint is binding. The fare component for leg i of product j is then given by its OD fare, minus the shadow prices of all legs except leg i. The intuition behind is that this shows the fare minus the revenue it displaced on other legs. There is no mathematical ground for this heuristic, but it is typically used in practice and is found to work well. However, this method also has drawbacks and may overvalue Equation (4.2).

To illustrate this, consider the example of Section 4.2.5 again. Assume that the shadow prices for flights $i = 1$ and $i = 2$, which we denote with s_1 and s_2, are both 0. This means that the fare component on flight $i = 1$ is equal to $\$1069 - s_2 = \$1069$, while the fare component for flight $i = 2$ is equal to $\$1069 - s_1 = \$1069$. In this example, this heuristic suffers from the same problem illustrated above: it overestimates the network value function. In practice, one crude heuristic that is used by scaling the fare components linearly so they sum to its OD fare. This heuristic is another approach that is not supported by any science.

4.4 Optimisation

In this section, we will inspect the question: **how do we determine the optimal policy?** This may seem easy: in fact, we have given the optimisation technique in Equation (4.2). It therefore may seem like it is plug and play of the demand and fare estimates. However, the practice of RM is different. In the following sections, we expose several shortcomings that make the problem much more difficult than it initially seems.

In Section 4.4.1, we highlight the inability of the RM system to frequently reoptimize, which is often said to be possible (and happen) in the theory. This is followed by Section 4.4.2, in which we expose how a lot of valuable data is lost in an attempt to store less information. In Section 4.4.3 we discuss how itineraries with a partner-airline component are optimized. One of the approximations that are being made is to solve the RM problem as a one-way problem. This is exposed in Section 4.4.4. The culture of focusing on individual legs, rather than the network, is investigated in Section 4.4.5. Analysts that override the system and the system unable to accurate read these strategies are introduced in Section 4.4.6. Optimisation including groups (or rather, excluding groups) is discussed in Section 4.2.3. The fixed denied boarding cost is a limitation of the system and is discussed in Section 4.4.8.

4.4.1 *Infrequent optimisation*

It is typically thought that optimization techniques, such as the leg-level dynamic programming or EMSR methods, are fast to find a solution. While the actual calculation of these optimization techniques is fast, one must not forget the data retrieval. All formulations need (at the bare minimum) demand estimates and fare estimates. Demand estimates are often stored in a (slow) database. Analysts often move demand up and down, based on their expertise. Not every RM system stores these actual demand forecasts; often, only the system generated forecast is stored. The system then needs to search for all user influences that are applicable for that product.

In practice, most influences in forecasting are done on a range of departure dates, often on a cabin-level. Naturally, the best practice indicates that demand influences should be surgical: for a given OD, for a single departure date, class, time of day, DCP and connection (if applicable). In practice, analysts often look at (multiple) POS, with a large number of ODs. This make specific demand influences almost impossible as there simply is no time to do so. In this case, it can be argued whether a (broad) demand influence even should be applied, but this is related to culture in the company.

Consider a demand influence for the AMS-SYD OD for January 1 - January 15 (with no restriction on POS / class / time of day / connection / etc), and one specific AMS-DXB flight is to be optimized. Before a demand estimate for the optimization is obtained, the system needs to iterate over all demand influences and check if the influence for this AMS-DXB is applicable. We have seen airlines that have over

45.000 of these such influences at the same in the system. Consider an airline that operates 500 flights a day, and flights up to 300 days before departure are for sale. Then this involves $500 * 300 = 150.000$ *for-loops* searching over 45.000 elements. This lookup alone can take hours.

Therefore, in practice, not every flight is reoptimized daily, and therefore, updating forecasts daily has little benefit. Typically, flights with departure dates in the near future (typically up to 60 or 90 days) are reoptimized on a daily basis, while flights in further in the future are only optimized every few days. One may argue that this isn't a bad approach, since flights in the distant future are yet to enter their booking curves, so correct availability is not as important as flights that are actively being booked. The problem with this approach, however, is there is a wide array of POS/OD traversing the same flight, and these do exhibit different booking curves. For example, there are POS that book very early, and thus it is important that the flights(s) that it covers are optimized frequently early on, to ensure proper availability for this POS/OD. However, the way optimization is scheduled in practice does not comply with this phenomenon. This may cause suboptimal availability.

4.4.2 *Time approximiation*

When using a dynamic programming approach to optimization, time is discretized in such a way that the probability of having more than one arrival at a given point of time is smaller than some chosen ϵ. Especially in DCPs (data collection points) with a big demand forecast, the number of time units is likely to outweigh the number of days in that DCP. Therefore, we end up with intra-day value function estimates (and, as a result, bid prices). Ideally, the vectors of value functions for each time unit t are used. However, in practice, after optimization, each day is assigned one such vector.

Suppose for example, that by dividing a time frame, we have estimates for $t = 43.1, 42.9, 42.7, 42.5, 42.3, 42.1, 41.9, 41.7, 41.5....$ It is then not immediately clear how an estimate for $t = 42$ is found. There are multiple ways, we may aggregate over values which round to $t = 42$ (which includes $t = 43.1, 42.9, 42.7, 42.5, 42.3, 42.1, 41.9, 41.7, 41.5)$. Alternatively, we may average over values whose floor is 42 ($t = 42.9, 42.7, 42.5, 42.3, 42.1$). Or, find the t closest to $t = 42$, in this case 41.9. Out of the heuristics introduced above, the last one is typically used in practice. This means that a lot of vectors of time t are discarded (note that because of the recursion, we have no option but to calculate them), and as a result, less information is used than is available to the system. It is typically argued that this is done to save storage space, but all heuristics shown above can be calculated before the vectors are discarded.

4.4.3 *Codesharing*

Many airlines have agreements with other airlines to offer connections beyond their own network. Airlines engage in these contracts as it dramatically increases the itineraries it can offer to their customers. For example, consider an airline that only offers a handful of flights to the United States of America: for instance, only to New York and Los Angeles. The airline realizes there is demand for other cities in this country, say, Boston, Chicago and Seattle, so it may approach an airline that flies from New York to Boston and Chicago, and an airline that flies from Los Angeles to Seattle. If the airline is successful in finding a partner, this partner will typically assign a flight number to this airline that it may use to sell this connection. This is why this partnership is called *codesharing*: the airline offers an itinerary with a flight number that resembles a typical flight number, but is operated by its partner. The benefit for both airlines is clear: the main airline is able to offer more connections, while its partner is given some share of the revenue. In practice, codesharing flights is done flight-specific. For example, suppose the airline's flight arrives in Los Angeles at 2PM. It then only approaches its partner for flights between say, 3 and 5PM to Seattle. Flights before 2PM and after 5PM are typically not included in a codesharing agreement, as it does not meet the very objective of codesharing: offering attractive itineraries to customers.

This does, however, pose an interesting question to demand forecasting. For example, consider the airline mentioned above that sold the itinerary AMS-LAX-SEA. It only operates the AMS-JFK flight themselves, while the LAX-SEA flight is operated by its partner. In Section 4.2.6 we discussed the challenges of the OD builder. This itinerary poses a similar difficulty: what is the OD offered? It may be tempting to assign the AMS-SEA OD, and use this in the optimisation. However, the difficulty here is two-fold: first, the number of data points for codeshare itineraries is generally relatively low, which makes forecasting difficult. Second, the crux of optimisation is deciding whether to accept a unit of capacity. However, because the LAX-SEA flight is operated by its partner, it does not have control over this process.

In practice, airlines make a distinction between the part of the OD it operates themselves, and is operated by others. This is typically defined as the online and offline OD. The *online OD is the part that is operated by the host airline*, while the *offline OD is the OD that is operated by both host and partner airline*. Only the online OD is operates themselves is used in optimisation. In the example above, the AMS-

LAX-SEA itinerary's online OD is AMS-LAX while its offline OD is AMS-LAX-SEA.

This has two implications from an optimisation perspective. First, the demand input λ for the product AMS-LAX is a combination of "true" AMS-LAX demand, and demand beyond LAX, as long as the online OD is AMS-LAX. Second, the fare in the optimisation is not immediately clear. The fare for the offline OD AMS-SEA is simply the fare paid, while the online OD AMS-LAX fare is typically prorated by distance, despite what airlines have contractually agreed. This introduces another challenge: this may cause a great spread in fares for the same product j: a AMS-SEA fare offered by the airline may be very similar to its AMS-LAX fare, but after proration of the former these are no longer similar.

4.4.4 *One way revenue management*

In Section 4.2.5, the concept of a network date was introduced. Equation (4.2) sums over value functions of all N flights for a specific network departure date. A direct result of this is that, in practice, revenue is maximized on a one-way basis. Consider an itinerary departing Amsterdam for London on the early morning of July 1, returning in the evening of July 2. This itinerary is split to two one-way parts: Amsterdam - London is optimized for a network departure date July 1, while the London - Amsterdam portion of this itinerary is used for the optimization of network departure date July 2. By doing this, we lose the effect of this being such a short trip. For example, even if it is known that this customer is a business traveller, and thus is typically thought to have a higher willingness to pay, the system is not able to exploit this information.

There are more implications. For example, now consider a passenger that books the same flight and booking class on July 1, but returns on July 10. It is much more likely that this customer is a tourist that is spending a holiday in the United Kingdom, and its trip purpose is very different from the passenger identified above. However, the RM system will aggregate over both these passengers to calculate and estimate demand, despite having very different return dates (and different trip purposes).

Another implication is that the system is unable to make a distinction between a "true" one-way itinerary, and a "derived" one-way itinerary. For example, consider a customer purchasing a one-way on July 1, with the same booking class as the passengers described above. Since this itinerary is a "true" one-way ticket, its "derived" one-way itinerary is the same. This means that this ticket is also included in the demand estimation. One-way tickets often are relatively expensive (for the same booking class), so including these to the same aggregate means that the (aggregated, product) fare estimate is skewed too.

Therefore, the system limitation is two-fold. First, it is impossible to determine availability for round-trips. Second, not making distinction between "true" and "derived" one-way itineraries means incorrect demand and fare estimates.

4.4.5 *Leg / Load factor culture*

In Section 4.3.2 we briefly introduced the revenue accounting department. The challenge airlines face is that they sells ODs, while they operate legs. Therefore, all costs it incurs (fuel, meals, crew cost) are leg based, while the revenue it generates is OD based. The revenue accounting department is tasked to, among other things, determine how much revenue is generated and how flights perform financially.

This inherently creates a problem. Airlines that use a network RM problem should have placed the performance of the network at the heart of everything they do, but in practice, the performance of individual flights is often prioritized. In practice, most airlines have daily reports that show load factor and yield information, that are sent to senior management. Load factor in particular is an element that is often discussed and prioritized in the department. This creates a focus on both leg and load factor-thinking that is hurtful to the performance of the airline's RM process. The concept of making more money with more empty seats is something that seems unbelievable to most employees. This creates a culture that is focused more on "getting bums on seats" than optimizing revenue.

This poses risks to the RM system: after all, if the objective is to maximize load factor, rather than revenue, it is beneficial to underforecast. Underforecasting lowers the value function and opens up availability. Increased availability for lower classes means recording more bookings (and thus achieving the (incorrect) objective). This, in turn, means larger forecasts for these lower classes, and reduced forecasts for higher classes. This process is repeated over time, and in the long run, this greatly affects revenues. This study is known as the spiral-down effect and has been studied extensively. This was briefly discussed in Chapter 2.

4.4.6 *Segment limits*

Intuitively, it is clear that managing a flight is much easier than managing ODs that traverse this flight. In practice, a flight may be utilized by tens of ODs. Each of these ODs have forecasts that depend on the class, DCP, travel time and more. Therefore, for a flight with a capacity of 400 it is not unusual to have more forecasts than units of capacity.

Now suppose that the availability for a flight is perceived "wrong". A revenue manager now has two options: he or she can talk to multiple demand analysts and ask them if their forecast is "correct" (we avoid a discussion on what "correct" is: there is no true definition of a "correct" forecast), or alternatively, ask a single flight analyst to adjust availability. If the manager opts for the first approach, the input will be fixed, but it may be days before the availability shows what is deemed optimal. If the manager decides to go for the second approach, the output will be fixed, but the problem is solved within seconds. Therefore, in practice, most managers will opt for the second approach. This approach is known as setting *segment limits*.

The availability, the booking classes for sale, expressed in terms of the letters of the alphabet, is the result of multiple products, each with very different fares. The issue with this approach is clear: closing booking classes in effect assumes that all these products have a same fare for every booking class. If this was in fact the case, this approach may not be so detrimental. In practice, however, there are great differences in fare for the same booking class. For example, consider the OD AMS-DXB. The fare for POS AE for T class is $700, while the fare for POS NL is $220 for that same booking class. Should the analyst decide to close T class, it will now reject a fare of $220 for Dutch passengers, but at the same time $700 worth of revenue from AE customers based is denied. This shows the danger of segment limits. Every RM system allows for these limits to be set. It is the result of legacy RM. In practice, segment limits are rampant.

From a system's perspective, these segment limits pose another danger. In practice, most RM systems only rely on OD availability, and do not have the technical capability to read segment limits. This means that the system does not recognize that this product is unable to be sold as a result of the segment limit, so this demand is not unconstrained. This, in turn, leads to underforecasting, and consequently the same spiral down effect as discussed in Section 4.4.5.

4.4.7 *Groups*

Most RM systems do not incorporate group requests in their optimisation. In Section 4.2.3, we have discussed that most systems do not forecast, and it is up to the analyst to evaluate the request. One of the reasons groups are not incorporated in Equation (4.1) is that this will introduce substantially more state transitions. In the formulation as is, there are only two state transitions: x or $x-1$ in the next time unit. Should groups be incorporated, further transitions to $x-2, x-3, .., 0$

need to be estimated and included in the calculation. The problem is two-fold. First, estimating the probability of these transitions will be difficult at best, impossible at worst, because of the high variance of group demand discussed in Section 4.2.3. Second, this will slow down the calculation of this newly-defined equation. This would mean less frequent optimisation, and this topic was discussed in Section 4.4.1.

Instead, most RM systems rely on heuristics to determine whether a group request is accepted. For individual requests, the fare is evaluated against the bid price, which is the opportunity cost of that seat that is for sale. In practice, one heuristic that is used that the total group fare is evaluated against the sum of the bid prices.

As discussed in Section 4.2.3, group demand behavior differs from individual demand in various ways, but arguably the biggest difference is its much higher cancellation rate, the direct result of very lax cancellation policies. For the airline, this means that overbooking is even more important, so as not to lose revenue as a result of capacity being released by groups cancelling. As the behavior of group demand is different, it calls for different methods of forecasting. In practice, we find that forecasting through machine learning techniques that take into account the agent type, its history, cancellation rules, contractual obligations are often much more accurate in predicting cancellations than traditional statistical methods. One limitation of RM systems, including those that feature group booking modules, is that forecasting is done in the same way as it is done for individuals.

4.4.8 *Denied boarding cost*

When airlines engage in overbooking in practice, analysts set a so-called *denied boarding cost*. This is often a fixed cost, and is used to determine how many seats one should overbook. This cost is independent of the products sold for that flight. Suppose, for example, a flight from AMS-JFK. Since this flight departs a EU-member state, airlines are liable to pay hefty compensation (€600) if a customer is involuntarily denied boarding. Now consider two flights, with two products for sale and capacity of 10 seats: the first flight is sold to five customers paying €750, while five others paid €400. The second flight were solely sold to ten passengers paying €400 each. In the first instance, it makes sense to overbook: after all, there are passengers that paid in excess of the denied boarding compensation. On the other hand, it doesn't make sense to overbook the second flight: an extra revenue of €400 is gained, but the airline is due €600, causing a net loss. Clearly, an optimal overbooking policy should take into account the current bookings of the flight.

One may argue that, in absence of loss of goodwill, it is optimal to keep accepting passengers until there are no more passengers that paid more than the denied boarding compensation. In the example above, consider the first flight. Suppose the flight is currently four seats overbooked. It may be worthwhile to overbook one extra seat at €750: this will generate €750 while the denied boarding compensation at €600 means the net profit to the airline is €100. If the overbooking limit is increased even further, this net profit becomes negative: -€200.

In practice, RM systems takes a fixed value for the denied boarding cost. There are systems that allow for denied boarding cost that depend on the number of bookings, but do not allow cost to depend on current bookings and therefore revenue received.

4.5 PRICING DISTRIBUTION

In this section, we will answer the last question we defined: **how do we ensure these fares are available to the customer?** So far, we have exposed issues in finding the optimal policy (and with that the demand and pricing inputs). Suppose now that we have successfully solved these problems. It is now important that these fares are available to the end-user. This process is called the *distribution*. In the following sections, we will introduce major limitations of distribution and discuss how these things may evolve in the future.

To understand the difficulties of distribution, we provide a brief overview of the global distribution system (GDS), which is at the heart of the process of distribution i. The limitations of current-day distribution are the direct result of the evolution of the GDS. These are discussed in Section 4.5.1. Next, we introduce the concept of IATA's New Distribution Capability (NDC). This is often referred to as the problem solver to the legacy issue of the GDS, but after an introduction we will show that this may create even more challenges than the industry faces today. This is investigated in Section 4.5.2.1.

4.5.1 *A brief history of time*

American Airlines faced a problem in the late 1950s, early 1960s. The number of passengers was at an all-time high, yet there was significant manual effort that went into selling a ticket to its customers. Partnering with IBM, it created the first reservation system of the industry: Semi Automated Business Research Environment (SABRE). At its time powered by a cluster of two of the most powerful computers in the world, it functions included look up of passenger reservations, and a first form of inventory control. This consequences of this development, that took over six years to complete, is still felt by the industry today.

It combined the functions of what we now consider a reservation system (storage of all reservations and cancellations) and RM system (maximizing revenue), as well as provided analytics capability to its management. Being developed in the 1960s, decades before the introduction of the internet for non-military, it meant that this was a system only accessible by the airline: it was an intranet, and was without any communication capability.

This proved to be a painpoint for SABRE. After all, the direct consequence of this system being an intranet, was that it was not accessible to outsiders. This made it a hassle for a travel agent: it had to call the airline to check for a schedule, check for pricing, check for availability, having the the airline's representative price a fare. It then had to relay this back to the customer. After the customer decides, it had to call back to the airline, check if the availability and fare was unchanged, and book a ticket. If the availability and pricing had changed, this process had to be repeated. This cumbersome process called the need for travel agents having direct access to SABRE. This posed two problems. First, the system was never designed to be accessed from outside the company. Second, travel agents should not be able to access the analytics part of SABRE.

We denote this internal, airline-facing system of SABRE, throughout as $SABRE_i$, with *i* for inward-facing.

4.5.1.1 *Further automation: Computerized Reservation System*

American Airlines, with its SABRE system realized this shortcoming and let travel agents access part of this system. The development of this system changed the industry, as it was the introduction of the *computer reservation system (CRS)*. We note this as $SABRE_o$, with the *o* indicating that this system is now accessible from the outside. Travel agents could see schedule, fares and availability: this greatly reduced the back-and-forth it was used to. SABRE also allowed other airlines to share their schedule, fare and availability data with them. This meant that travel agents were not only able to book American Airlines flights, but also others. Research [152] shows that there was a bias against other airlines, which often appeared lower in the ranking than American Airlines itineraries, despite being more attractive. This, in turn, caused the development of similar system by other airlines. Delta Airlines and United Airlines quickly followed. Airlines made money by charging travel agents by search query, and other airlines paid a fee for every segment that was stored in another airline's reservation system. This historical event is important to note, as this turned out to be the birth of the GDS fee, that is often discussed by airlines. Airlines realized that there was a lot of money to be made from transactions made in their *CRS*, and this caused them to develop their own techniques and processes, which in turn resulted in proprietary systems. For example, American Airline's SABRE system was not compatible with United Airline's APOLLO.

4.5.1.2 *Further automation and extension: Global Distribution System*

In the 1980s, SABRE was sold by American. Other systems were sold by airlines too, and the new companies that sold (the usage of) these reservation systems, were now separate entities and, in theory, no longer biased toward a single airline. Technological advances were present, but all of these were an evolution of the reservation system, which in turn was the evolution of SABRE. The role of advanced system, dubbed *global distribution system (GDS)* was similar: all airlines connected to it had its fares and schedules shown, as well as the availability for the booking classes. This posed the question: if these systems are no longer biased by an airline, like for example $SABRE_o$ was, in what order are itineraries presented to the end user? Technical limitations made it impossible for airlines to "bid" to be at the top of the search results, as it is nowadays possible to do in for example Booking.com or Google. Instead, travel time and schedule were the parameters that were used to order results [153]. As this was mutually agreed, all airlines were forced to have these elements present in the GDS. This also turned out to be an important historical event.

However, it didn't end with just a GDS. Up to this point, customers would need to approach a travel agent to make a booking, which was done through the GDS. With the rise of the internet, it was made possible to purchase tickets online. However, not every platform wanted to integrate with each GDS separately, as connecting to every GDS took significant effort - after all, every GDS had their own input and language. This caused the rise of GDS aggregators. These would be the interface between the GDSs and the different platforms.

4.5.2 *Ongoing developments*

Through a brief history of the development of the GDS, we have made it clear what caused the challenges of current-day distribution. In summary, we have identified the following events that have caused consequences that are still felt by the industry today:

- Systems that were not designed for outside communication (American Airlines, $SABRE_i$);
- Future development of $SABRE_i$ being a band-aid approach, newly identified needs built on top of a system not designed for it;

- Biasing results toward own airline, causing the rise of multiple, competing GDSs (American Airlines, $SABRE_o$);
- Hard-coded search parameters by the industry - travel time, schedule and fare.

In the next sections, we will introduce IATA's answer to these challenges: their new distribution capability. We provide a brief overview of its history, provide a brief of current usage, and provide a list of challenges that NDC introduces.

4.5.2.1 *IATA's New Distribution Capability*

IATA, the International Air Transport Association, sets out standards for the aviation industry. Realizing the shortcomings of the GDS, they launched the *New Distribution Capability (NDC)* initiative in 2012. Up to this point, requests between and within airlines were done using EDIFACT (Electronic Data Interchange for Administration, Commerce and Transport). This standard, developed before the rise of internet, independent of the industry, was no longer suitable for complexities the industry faced. One of the technological advancements was the replacement of EDIFACT: initially, this was done through XML. However, as years have progressed, these requests are now JSON. While it is important to understand the new possibilities and opportunities that NDC offers, this technological advance is critical as it offers a much faster and efficient way of communication.

In Section 4.5.2, we discussed that several parameters were hard-coded in the GDS. As airlines grew and explored different ways to differentiate themselves from others, new services were being offered. For example, airlines like Virgin Atlantic and Emirates offers a chauffeur drive from and to the airport, making the travel experience a seamless experience. However, agents were not able to filter or order based on these amenities. Other airlines have started to unbundle their fare offering, aiming to personalize the offering to its customers. *Unbundling* means that a customer pays a basic fare for a seat, but other services such as a checked bag, lounge access, or miles come at an additional fee. Through the traditional GDS this was impossible. In an ideal world, NDC will fix this problem. Every request will receive a personalized offer. This is referred to as *dynamic offer generation (DOG)*.

While the purpose of NDC was to standardize the language used in the industry, through initially XML and, later on, JSON requests, current adoption is far from standard. Every airline has its own API it uses to interface connectivity between API and the newly found aggregators - each with different functionality. After all, every airline wants to set themselves apart and they realize this can be achieved through a superior API. As Misquitta [154] points out, even within the three big airline alliances, there are differences. For example, within OneWorld, American Airlines used a different approach than the IAG group. This shows the challenge that NDC has created: while its objective was to standardize connectivity, the freedom it creates by allowing multiple standards has only made things more complicated. IATA has responded and said they would certify NDC connections, but there is little clarity what this constitutes and what the "right" approach is.

4.5.2.2 *Future and challenges*

The challenges identified are as follows: standardization of NDC version, integration with other airlines and ticketing.

Firstly, the industry should agree on what NDC version to use. As Misquitta [154] points out, a new version is released every six months. Therefore, it is important that the industry agrees on what standard to use. In practice, however, it is not unthinkable that there will be many different versions that will be used: while different versions give airlines more choice in what works best for them, it does also introduce capability issues. Secondly, it is important that airlines are able to integrate with one another. As more and more airlines work together through different levels of partnerships, NDC can only be successful if connections can be integrated. This will prove to be difficult: compare this to the current GDS landscape - there are "only" three GDSs which airlines need to integrate. In the future, as a result of both airlines wanting to set themselves apart and the variety of NDC versions available, this may cause severe challenges. In an extreme case, where every airline has their own NDC setup, we are back at the time before the CRS: every airline has their own interface. Thirdly, it is no longer clear who is the ticket issuing carrier. Consider an itinerary that is a combination of a carrier that is connected to GDS, one that is connected via NDC, and one that is connected directly. One of the advantages of the GDS that it stores the PNRs that are issued by that airline. In this case, there may be need for master PNR database, that links the three different portions of the aforementioned ticket.

4.6 DISCUSSION

In this chapter, we have discussed both technical limitations and challenges with culture that limit the effectiveness of RM. We can group these limitations in a number of different categories: data storage and computation time, approximations with merit, approximations without merit, and culture.

Data storage and computation
To limit the data (vectors by time slice) stored, RM systems dispose data and only store one vector per day before departure. This was introduced in 4.4.2. While this approximation is likely not to cause any changes in the availability of booking classes, it does warrant research to study what the best aggregation technique is. On the other hand, on critical flights, especially toward the last few stages of the booking curve, this approximation may have a negative impact on revenue.

In Section 4.4.1, computation time was discussed. What looks like a simple dynamic programming formulation, is preceded by data retrieval. Finding demand estimates λ can take hours alone, and therefore the availability is recalculated not as frequently as the airline would like. Should an airline experience thousands of forecast overrides, it should critically review both the RM system and the company culture. Reducing these demand influences by themselves may make it possible to reoptimize more frequently.

Approximations with merit
Several approximations that have some level of merit were identified. In Section 4.4.7, it was discussed the policy for accepting groups is given by taking the sum of bid prices. While this violates the underlying Markov Chain, it is the only intuitive approximation available. It should be added, however, that in practice analysts may override this, and this, as a result would violate this being an approximation with merit. The concept of one-way RM was introduced in Section 4.4.4.The curse of dimensionality of adding return sectors in defining products j make this a reasonable approximation. It is then important that appropriate fare rules are set so the airline is able to have a quasi-control over return RM. The challenges poses of codesharing were illustrated in Section 4.4.3. The dangers arise from mixing online and offline ODs and fare evaluation, but this approximation is still reasonable. It is not entire clear how to best split network departure dates, which was discussed in Section 4.2.5. Similarly, the difficulty of defining ODs from a given itinerary is discussed in Section 4.2.6. This can be seen as a reasonable approximation, as long as the business rules defined to split ODs are aligned with the fare rules (which often is not the case).

Approximations without merit
Section 4.2.1 illustrated how the RM system decides whether demand is to be unconstrained. Clearly, only looking at the last day of the DCP to determine whether a class is closed is a very crude approximation, yet it is often used in practice. It can be argued why the notion of DCPs is used in the first place. After all, if days are not grouped, it should be easy to determine whether a class is closed for sale or not. However, even there intra-day availability changes create the same challenge. In

Section 4.2.2, it was discussed how the RM system uses a single set of DCPs for all its ODs and flights across the network. There is absolutely no need to have consistent DCPs across the network, since it is always possible to align vectors with a different number of time units. It is likely that RM system vendors use static DCPs to make data lookup faster. In Section 4.2.3, group demand was reviewed. It was mentioned that group demand poses a significant risk to the airline, because of very lenient cancellation and change penalties. Despite this, airlines continue to offer these policies to travel agent that actively abuse this goodwill. There is an easy fix for this: make these policies more restrictive, similar to the conditions offered to individual customers. Lastly, in Section 4.2.4 cancellation estimation was discussed. This is another element in the system that provides no merit: in the ideal case, gross demand and cancellations are modelled subsequently optimized over. One approach to modelling cancellations was introduced in Chapter 7 and a optimisation heuristic was proposed in Chapter 9. The way OD fares are leg-prorated and subsequently used in the optimisation, introduced in Section 4.3.2, is another approximation that does not have mathematical foundation, despite being the most popular heuristic used in practice. Lastly, the differences in final revenue received by the airline, introduced in Section 4.3.1, illustrate the dangers of calculating fares used in optimisation. In practice, this is ignored and this is clearly dangerous.

Culture

Lastly, culture in the RM department may hinder revenue performance. Up to this day, there exists a "load factor culture" in many airlines. This was discussed in Section 4.4.5. While academia have long agreed that (approximate) network RM outperforms leg RM techniques, and therefore the focus should be on network performance, it may be argued that it is "easier" to look at load factors. In some airlines, there is no distinction between "revenue" and "non-revenue" load factor: the latter includes staff travel and involuntary upgrades. Clearly, these should be excluded in reporting as these do not generate revenue (hence, "non-revenue"). In Section 4.4.6, segment limits were discussed. This is related to the "load factor" culture, and may cause substantial revenue loss as the fare values differ widely across different POS and ODs. This is followed by Section 4.4.8, in which the concept of *denied boarding cost* was introduced and it became evident that this is a static, leg-level based value set by an analyst. Two problems were found: first, this introduces subjectivity, and second, this cost should be dependent on current bookings. Lastly, the industry cannot seem to agree on what the best way forward in terms of distribution is. This is covered in Section 4.5: while IATA's NDC was introduced to solve the challenges airlines faced, because of the flexibility it offers, and diversity of versions offered, it may cause more problems than it intends to solve.

4.7 CONCLUSION

1. Most methods proposed in the literature cannot be put in practice for different reasons. In Chapter 1 we saw that one reason was simplifying assumptions, in this chapter, we observe system limitations.
2. Since inventory is a daily measure while DCP's span multiple days, it is not clear when a class is "closed" for sale.
3. In the most frequently used RM solutions [122] DCP's are used across the network. Using the same DCP for different OD's is challenging, since different OD's have different booking curves.
4. Group bookings violate the assumption of having at most one arrival within a (discretized) time interval, and are often ignored.
5. Itineraries with connections that span multiple days have to be split, causing either incorrect demand forecasts or sub-optimal optimisation.
6. Determining an OD-pair from an itinerary is not straightforward and requires manual inputs.
7. Fares after fees and charges across different distribution channels are different but are equally used.
8. Determining the value of an OD-pair on an individual leg is difficult and current methods overestimate the value function.
9. Even dynamic programming methods with a single-dimension state space are too slow to run daily, causing infrequent optimisation.
10. While time is discretized in the optimisation method, this data is disregarded in the output.
11. Optimisation of OD-pairs that include codesharing is difficult because of differences in the fare used for optimisation and actual revenue received.
12. Optimisation is done on a one-way level while the airline most frequently sells round-trips.
13. Denied boarding cost is often static, and not dependent on the products already sold.
14. In practice, many airlines have a "load factor" culture, rather than a "revenue" culture.

Part III

THE REAL THEORY OF REVENUE MANAGEMENT

5 UNCONSTRAINING DEMAND USING GAUSSIAN PROCESSES

Abstract

One of the key challenges in RM is unconstraining demand data. Existing state of the art single-class unconstraining methods make restrictive assumptions about the form of the underlying demand and can perform poorly when applied to data which breaks these assumptions. In this chapter, we propose an unconstraining method that uses Gaussian process (GP) regression. We develop a novel GP model by constructing and implementing a new non-stationary covariance function for the GP which enables it to learn and extrapolate the underlying demand trend. We show that this method can cope with important features of realistic demand data, including nonlinear demand trends, variations in total demand, lengthy periods of constraining, non-exponential inter-arrival times, and discontinuities/changepoints in demand data. In all such circumstances, our results indicate that GPs outperform existing single-class unconstraining methods.

This chapter is based on [19].

5.1 INTRODUCTION

5.1.1 *Demand unconstraining for revenue management*

Airlines commonly set booking limits on the number of cheaper fare-classes that can be purchased, or make cheaper fare-classes unavailable for booking at certain times in an attempt to divert some of that demand to the more expensive tickets still available. While a fare-class on a given flight route is available for booking, the demand for that 'product', at that price, is accurately captured by its total recorded bookings. However, once the product has been unavailable for booking for a period of time, recorded bookings no longer capture true demand, and the demand data is said to be 'constrained' or 'censored'.

Practices which constrain demand data pose a big challenge for successful RM. This is because many important decisions, including setting ticket prices, making changes to an airline's flight network, adding or removing capacity on a certain route, and many others, are all heavily dependent on accurate historical demand data. Moreover, precisely those decisions regarding which fare-classes to make unavailable (and for what periods of time) themselves depend on accurate demand data.

Since the mid-1990s, researchers have been studying ways to manage the constrained demand problem, and the proposed approaches fall under the banner of 'unconstraining' methods. Broadly speaking, there are two types of product-model for which unconstraining methods are developed: single-class models which assume that demand for each fare-class is independent of the availability of (and demand for) all other fare-classes; and dependent demand models, where demand for a given fare-class on a given flight depends on the availability of (and demand for) other fare-classes on the same flight (or even other flights as well).

While assuming dependent demand might be theoretically, as well as practically appealing, multi-class methods have not been widely adopted in practice [66]. The resulting consumer-choice based unconstraining methods can be much more complicated and expensive to incorporate at scale into airline RM systems, and rely on methodologies for estimating choice parameters and arrival rates which are as-of-yet ineffective, see Guo et al [66].

For these reasons, we focus in this chapter on the single-class unconstraining problem, which can be stated as follows: *given a demand curve which becomes constrained at some point, how do we accurately predict what demand would have been had no constraining occurred?* We illustrate this problem in Figure 5.1, which highlights the difference between constrained and true demand as a consequence of an imposed booking limit.

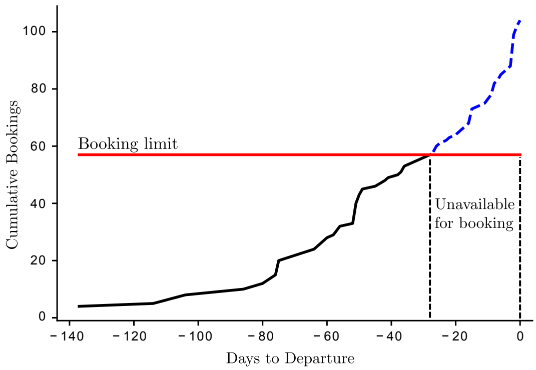


Figure 5.1: Schematic illustrating the single-class unconstraining problem.

Figure 5.1 displays a demand curve. This demand curve shown represents cumulative bookings for a particular fare-class on a given flight. Observed bookings accurately capture true demand until the fare class becomes unavailable for booking due to an imposed booking limit, after which time the observed demand remains constant at the booking limit, even though the true demand exceeds that limit.

5.1.2 *Mathematical formulation of demand unconstraining*

For clarity and consistency across the different unconstraining methods, we formulate the unconstraining problem in general terms. Let $\mathcal{D}$ be the full set of demand observations under consideration (which may or may not be ordered, depending on the unconstraining method in question). Next, define:

1. *True demand*: demand observations, made when demand was not constrained in any way;
2. *Constrained demand*: demand observations made in the presence of a constraint;
3. *Unconstrained demand* refers exclusively to the *output produced by an unconstraining method*, that is, approximations of what the constrained demand observations would have been had there been no constraining.

We define $\mathcal{D}_T \subset \mathcal{D}$ as the subset of true demand observations, and $\mathcal{D}_C \subset \mathcal{D}$ as the subset of constrained demand observations, such that $\mathcal{D} = \mathcal{D}_T \cup \mathcal{D}_C$ and $\mathcal{D}_T \cap \mathcal{D}_C = \emptyset$. We define $\mathcal{D}_U$ as the set of unconstrained demand values, corresponding to the unconstrained approximations of the elements of $\mathcal{D}_C$, and define $\widehat{\mathcal{D}} = \mathcal{D}_T \cup \mathcal{D}_U$, the full set of demand values where the constrained observations have been replaced by their unconstrained approximations. The demand unconstraining problem can therefore be stated mathematically as: use the available demand observations $\mathcal{D} = \mathcal{D}_T \cup \mathcal{D}_C$ to estimate the unconstrained demand values $\mathcal{D}_U$ as accurately as possible.

5.1.3 *Gaussian processes*

Gaussian processes (GPs) provide a robust statistical basis for inferring underlying statistical models from observed data. GPs were first applied to time series analysis in the work of Wiener [155] in 1949, but it was not until the 1970s that a general theory of GP prediction was developed by O'Hagan and Kingman [156]. Since then, GPs have become a popular and very general framework for statistical modelling, and have been used to tackle a vast array of problems, including applications in machine learning, which can be found in, for example, Rasmussen [157], atmospheric modelling by Fuentes and Raftery [158], biochemical reactions by Gao et al. [159], and many others.

One of the applications for which the use of GPs has yet to be investigated, however, is unconstraining demand. In this chapter we propose and test a new GP model for use on this problem. We show that it outperforms state of the art unconstraining methods, coping much better with nonlinear and even discontinuous demand trends, variations in total demand, lengthy periods of constraining, and both exponential and non-exponential inter-arrival times.

The rest of this chapter is structured as follows. Gaussian processes are introduced: first, in Section 5.2, we briefly introduce the mathematics of GP regression; we then move on, in Section 5.3, to motivate and develop the details of our GP model for application to the single-class unconstraining problem. In Section 5.4, we describe three numerical experiments conducted to evaluate the performance of our proposed GP unconstraining method in comparison to state of the art alternatives using generated data that more accurately resembles real Emirates Airlines bookings data. In Section 5.5 we extend our proposed unconstraining method to handle demand trends which exhibit kinks and discontinuities, illustrating its performance on three scenario-inspired test cases. A discussion is given in Section 5.6. We close with conclusions in Section 5.7.

5.2 GAUSSIAN PROCESS REGRESSION

The general idea behind Gaussian Process regression is very intuitive. We start by assuming a prior Gaussian distribution over functions, and then restrict our distribution to include only those functions which make sense given the observed data. More formally, our goal is to infer some unobserved (latent) function f evaluated at a set of test inputs $X^* = \{x_1^*, \ldots, x_m^*\}$, using observed data $\mathbf{y} = \{y_1, \ldots, y_n\}$ at training points $X = \{x_1, \ldots, x_n\}$. Let $\mathbf{f}$ and $\mathbf{f}^*$ be vectors of unobserved function values at inputs X and X^* respectively, and let θ_c be a set of covariance hyperparameters for f.

The function $f(x)$ is a GP if, for every tuple x, $f(x)$ is a random variable, and for any finite set of points $\{x_1, x_2, \ldots, x_n\}$, the set $\{f(x_1), f(x_2), \ldots, f(x_n)\}$ has joint Gaussian distributions whose mean and covariance are defined by a mean function $m(x)$ and covariance function $k(x, x')$ evaluated at the points $\{x_1, x_2, \ldots, x_n\}$. We will assume throughout that the mean function is zero, as is standard in the literature. A covariance function takes the form of a kernel (or similarity) function mapping $x, x' \in X$ to $\mathbb{R}$, which specifies the covariance between the random variables $f(x)$ and $f(x')$, denoted as

$$\text{Cov}[f(x'), f(x)] = k(x, x').$$

Covariance functions are symmetric by definition and require that the covariance matrix $\mathbf{K}_{i,j} := k(x_i, x_j)$ of the points $\{x_1, x_2, \ldots, x_n\}$ must be Positive Semi-Definite (PSD) [157]. A covariance function is said to be 'stationary' if it is a function of $(x - x')$, making it invariant under translations in input space [160]. In contrast, when k is not a function of $(x - x')$ the covariance function is known as 'non-stationary'.

Assuming a GP prior implies a joint Gaussian distribution over $\mathbf{f}$ and $\mathbf{f}^*$, and it is known [157] that the conditional distribution of $\mathbf{f}^*$ given $\mathbf{f}$ is

$$\mathbf{f}^*|\mathbf{f}, X, X^*, \theta_c \sim \mathcal{N}(\mathbf{K}_*^\top \mathbf{K}^{-1}\mathbf{f},\ \mathbf{K}_{**} - \mathbf{K}_*^\top \mathbf{K}^{-1}\mathbf{K}_*), \tag{5.1}$$

where the covariance matrices $\mathbf{K}_{i,j} := k(x_i, x_j)$, $(\mathbf{K}_*)_{i,j} := k(x_i^*, x_j)$, and $(\mathbf{K}_{**})_{i,j} := k(x_i^*, x_j^*)$.

In general, the function f is considered to be a latent function, meaning that we do not observe the actual function values $\mathbf{f}$; rather we observe values $\mathbf{y}$ which are related to the true function values in a particular way. This relationship is defined by the observation model, that is, the likelihood of the observed values $p(\mathbf{y}|\mathbf{f}, X, \theta_l)$ where θ_l denotes the set of likelihood hyperparameters. The specific form of the likelihood depends on the process one is trying to model and will not be Gaussian in general.

Given our likelihood $p(\mathbf{y}|\mathbf{f}, X, \theta_l)$ and our prior $p(\mathbf{f}|X, \theta_c)$, we need to calculate the conditional posterior distribution $p(\mathbf{f}|\mathbf{y}, X, \theta)$, where $\theta = \theta_c \cup \theta_l$. Bayes' rule, the cornerstone of Bayesian inference, allows us to compute this as follows:

$$p(\mathbf{f}|\mathbf{y}, X, \theta) = \frac{p(\mathbf{y}|\mathbf{f}, X, \theta_l)p(\mathbf{f}|X, \theta_c)}{\int p(\mathbf{y}|\mathbf{f}, X, \theta_l)p(\mathbf{f}|X, \theta_c)\mathrm{d}\mathbf{f}}. \tag{5.2}$$

In cases where the likelihood (observation model) $p(\mathbf{y}|\mathbf{f}, X, \theta_l)$ is Gaussian, the marginal likelihood (the integral in the denominator) can be calculated exactly. In all other cases, the conditional posterior must be approximated. One standard approach is to construct a Gaussian approximation of the conditional posterior using the Laplace approximation [157], yielding

$$p(\mathbf{f}|\mathbf{y}, X, \theta) \approx \mathcal{N}(\hat{\mathbf{f}}, \Sigma^{-1}), \tag{5.3}$$

where the mode $\hat{\mathbf{f}} := \arg\max_{\mathbf{f}} p(\mathbf{f}|\mathbf{y}, X, \theta)$ and the precision matrix Σ is the Hessian of the negative log conditional posterior evaluated at the mode:

$$\Sigma = -\nabla\nabla \log p(\mathbf{f}|\mathbf{y}, X, \theta)|_{\mathbf{f}=\hat{\mathbf{f}}} = \mathbf{K}^{-1} + \mathbf{W}, \tag{5.4}$$

where $\mathbf{W}$ is the diagonal matrix with entries $W_{ii} = \nabla_{f_i}\nabla_{f_i} \log p(y_i|f_i, x_i, \theta_l)|_{f_i=\hat{f}_i}$. For details of calculating the mode $\hat{\mathbf{f}}$ and the precision matrix Σ see [157].

Finally, to obtain the posterior predictive distribution, we combine the conditional probability (5.1) with the conditional posterior (5.2) and marginalise out the latent function values $\mathbf{f}$:

$$p(\mathbf{f}^*|\mathbf{y}, X, X^*, \theta) = \int p(\mathbf{f}^*|\mathbf{f}, X, X^*, \theta_c) p(\mathbf{f}|\mathbf{y}, X, X^*, \theta) \mathrm{d}\mathbf{f}.$$

In cases where the conditional posterior has been approximated with the Laplace approximation (5.3)-(5.4), computing this integral gives

$$\mathbf{f}^*|\mathbf{y}, X, X^*, \theta \sim \mathcal{N}(\boldsymbol{\mu}_p, \mathbf{K}_p), \tag{5.5}$$

where the posterior mean and covariance are given by

$$\begin{aligned} \boldsymbol{\mu}_p &= \mathbf{K}_*^\top \nabla \log p(\mathbf{y}|\mathbf{f}, X, \theta_l)|_{\mathbf{f}=\hat{\mathbf{f}}}, \\ \mathbf{K}_p &= \mathbf{K}_{**} - \mathbf{K}_*^\top [\mathbf{K} + \mathbf{W}]^{-1} \mathbf{K}_*. \end{aligned} \tag{5.6}$$

Before we can make predictions using the posterior predictive distribution (5.5), we remove its dependence on θ by marginalising out the hyperparameters by integrating with respect to θ. To do this we need to compute

$$p(\mathbf{f}^*|\mathbf{y}, X, X^*) = \int_\theta p(\mathbf{f}^*|\mathbf{y}, X, X^*, \theta) p(\theta|\mathbf{y}) \mathrm{d}\theta. \tag{5.7}$$

We apply Bayes' rule to $p(\theta|\mathbf{y})$ to transform the integral in (5.7) into one in terms of the marginal likelihood $p(\mathbf{y}|\theta)$. The resulting equation is

$$p(\mathbf{f}^*|\mathbf{y}, X, X^*) = \frac{1}{Z} \int_\theta p(\mathbf{f}^*|\mathbf{y}, X, X^*, \theta) p(\mathbf{y}|\theta) p(\theta) \mathrm{d}\theta, \tag{5.8}$$

where $Z = \int_\theta p(\mathbf{y}|\theta) p(\theta) \mathrm{d}\theta$ is the marginalisation constant, and $p(\theta)$ is a prior distribution over our hyperparameters which must be specified. Garnett et al. [161] approximate this integral with Bayesian Monte Carlo techniques, while Saatçi et al. [162] recommend a simple quadratic approximation instead. We choose to implement the latter, approximating the integrals in (5.8) with sums such that

$$p(\mathbf{f}^*|\mathbf{y}, X, X^*) \approx \sum_{\theta_g} p(\mathbf{f}^*|\mathbf{y}, X, X^*, \theta_g) \left(\frac{p(\mathbf{y}|\theta_g)}{\sum_{\theta_g} p(\mathbf{y}|\theta_g)} \right),$$

where $\{\theta_g\}$ is a grid placed over a reasonable subspace of the GP hyperparameters, and where we have assumed a uniform prior probability mass at each grid point.

5.3 GAUSSIAN PROCESSES FOR UNCONSTRAINING DEMAND

5.3.1 *Problem setup*

The single-class unconstraining problem, from a time-series perspective, is equivalent to a short-term forecasting or extrapolation problem.

We propose using GP regression to learn the underlying booking trend for a particular flight from the bookings data up until the time it was made unavailable (i.e. from true demand observations), and to make predictions about what the true demand would have been thereafter.

Unlike double exponential smoothing (DES), which takes cumulative bookings as inputs and forecasts directly in 'cumulative-space', our aim is to model the underlying booking trend, and hence, we perform GP regression on daily bookings. We take the set of points X to be the days from when tickets were made available until the day they were constrained, and the observed data $\mathbf{y} = \mathcal{D}_T$ to be the set of observed daily bookings on these days. Since we are in 'daily-space', the constrained demand observations in $\mathcal{D}_C$ are all zeros, corresponding to inputs X^*, which are the remaining days before departure when booking was not possible. Once we have defined a suitable model, described in the remainder of this section, we follow the steps described in Section 5.2 to calculate the posterior predictive distribution $p(\mathbf{f}^*|\mathbf{y}, X, X^*)$ and forecast our daily unconstrained demand values to be the mean of this posterior predictive distribution. This approach is illustrated in Figure 5.2 below.

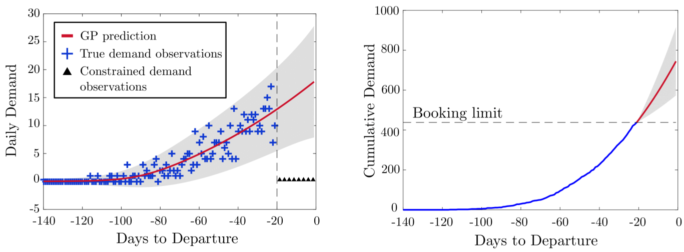


Figure 5.2: Illustration of GP regression for unconstraining demand. The figure on the left shows the mean prediction and the confidence interval produced by our GP method, based on the true demand observations. The dotted black line indicates when the booking limit was reached, and the red line beyond this point shows the GP's unconstrained approximations. The figure on the right shows in red the reconstruction of the cumulative demand curve over the constrained period using the daily demand values predicted with the GP.

5.3.2 *Motivation*

Multi-curve methods (using entire booking curves for unconstraining), though currently favoured by airlines, face a number of key challenges. Firstly, they require a significant amount of true historical demand data to perform well. Crucially, this must be data from flights which are assumed to have similar booking patterns and demand totals, since multi-curve methods all make the assumption that the demand data they use share a common underlying distribution. Since booking behaviours for flights vary with the month of departure, the weekday of departure, and even the time of departure, there needs to be a long history of accurate demand data for these methods to be accurately applied. Not only is this a problem for new flight routes which do not yet have sufficient demand history; it is also a potential problem for popular fare classes on peak-season. These flights are almost always constrained at some point before departure. For these flights, while there is a long history of recorded demand data, very little of this will be true demand (as is necessary). This problem is clearly avoided by single-curve methods, for which the only data needed to unconstrain a particular demand curve is the demand data from that curve prior to it being constrained.

A second problem for multi-curve methods is their inability to account properly for exogenous circumstances which change over time. Flight demand is affected by the strength of the economy, inflation and ticket prices, the relative strength of the origin and destination currencies, and many other factors. These are likely to vary over time, creating unaccounted-for variation in demand even among flights departing at the same time on the same weekday but in different months or years. Since multi-curve methods prioritise producing unconstrained demand estimates which are similar to past flights from different months and years, they cannot adequately take account of these exogenous effects. Single-curve methods, on the other hand, implicitly consider these by utilising only the trend in demand for that specific flight up until it was constrained.

For these reasons, we favour a single-curve approach. DES, the only other distinctly single-curve method, has a number of key limitations, the most important one being that it can only produce linear extrapolation. GPs, in contrast, have the ability to learn and extrapolate non-linear trends, which is an important advantage.

5.3.3 *Proposed model*

Our proposed GP model is based on the assumption that the flight bookings process (which can be thought of as an 'arrivals process') is best modelled as a Cox process (otherwise known as a doubly-stochastic Poisson process). This is a generalisation of the standard assumption made in RM [65] which models bookings as an inhomogeneous Poisson point process. In a Cox process, however, the time-dependent $\lambda(t)$ is itself a stochastic process (a Gaussian process in this case). To this end, we use a Poisson likelihood for our observation model $p(\mathbf{y}|\mathbf{f}, X, \theta_l)$, which is to say that we assume the observed bookings on day x_i to be a sample from a Poisson distribution. However, we cannot take $y_i \sim \text{Poiss}(f(x_i))$ since when $f(x_i) = 0$, this is not a distribution. A standard approach [163] is therefore to treat the number of bookings on day x_i as a sample from a Poisson distribution with rate $\lambda(x_i) = \log(1 + \mathrm{e}^{f(x_i)})$.

As mentioned in Section 5.2, we use a zero mean function for the GP prior, and our choice of covariance function is influenced by two key considerations. The first is that we are going to be using our posterior distribution for the purposes of extrapolation, since we are predicting what the demand trend would have been beyond the time at which it was constrained. The second is that we assume that the rate $\lambda(t)$ of the underlying inhomogeneous Poisson process is generally smooth (though we do not exclude the possibility of sudden, infrequent changes in the scale of (and/or trend in) demand, with which we deal explicitly in Section 5.5).

The extrapolation consideration is important since when performing GP regression using most stationary covariance functions, the posterior mean tends towards the prior mean as one moves further away from the observed data $\mathbf{y}$, making these covariance functions poor candidates for applications involving extrapolation. For better performance on extrapolation problems, Wilson et al. [164] propose a so-called spectral mixture covariance function, which uses a weighted product of multiple Gaussians in constructing the spectral density of a new stationary covariance function. However, while their results are impressive, their covariance function does not entirely avoid the 'mean problem' faced by other stationary covariance functions. Furthermore, its performance is highly sensitive to its hyperparameter initialisation, requiring a computationally expensive initialisation procedure in order to choose appropriate initial values.

5.3.4 *A non-stationary covariance function*

We propose a non-stationary covariance function for our GP model that does not suffer from the 'mean problem' faced by most stationary covariance functions. Further motivation for a non-stationary covariance function comes from considering the bookings process we are attempting to model. Consider an inhomogeneous Poisson process with rate $\lambda(t)$. Let $B(t)$ be a random variable representing the total number of bookings in the window from time 0 to time t, such that $\text{Var}[B(t)] = \mathbb{E}[B(t)] = \int_0^t \lambda(s)\mathrm{d}s$ is the variance of $B(t)$. For s such that $0 < s < t$, $B(s)$ and $B(t) - B(s)$ are independent and thus have a covariance of zero. This means we can write $\text{Cov}[B(s), B(t)] = \text{Cov}[B(s), B(s)] + \text{Cov}[B(s), B(t) - B(s)] = \text{Var}[B(s)]$, which shows that the covariance in 'cumulative-space' is non-stationary.

Our proposed covariance function is motivated by the polynomial covariance function [157]:

$$k(x, x') = \sigma^2(x^\top x' + c)^p, \tag{5.9}$$

where $\theta_c = \{\sigma, c\}$ are hyperparameters and the degree p is some specified positive integer. For a given degree, GP regression with this covariance function can be shown to be equivalent to Bayesian polynomial regression [157]. This serves as a sensible starting point for our model, since the 'smoothness' assumption on the underlying Poisson rate means that it can likely be well approximated with a polynomial.

5.3.5 *Automatic degree inference*

Since we do not know a priori what the degree p of the polynomial covariance function (5.9) will be, we propose a new covariance function that treats p as a hyperparameter as well, to be inferred from the data like σ and c. Therefore, our proposed covariance function is also of the form of (5.9), but where this time $\theta_c = \{\sigma, c, p\}$ are the covariance hyperparameters.

Once p is a hyperparameter, it cannot be restricted to integer values. Polynomial kernels with fractional degree are not unprecedented, however. Kernels of the form $(x^T x')^p$, for $0 < p < 1$, have been used before for facial recognition using the kernel PCA method [165]. Also, and Rossius at al. [166] discuss the use of kernel functions of the form $(x^T x' + 1)^p$ in Support Vector Machines, and the impact of non-integer values of p when $x^T x' < -1$, in which case the base raised to a non-integer power is negative. In both cases, however, p was considered to be a fixed (albeit non-integer) value and we propose to generalize the GP regression framework by letting p be a covariance function hyperparameter which is automatically inferred from the data. To the best of our knowledge, this has never been done before.

With our proposed covariance function, the covariance matrix becomes

$$\mathbf{K} = \sigma^2(\mathbf{x}\mathbf{x}^\top + c\mathbf{e}\mathbf{e}^\top)^{\diamond p},$$

where $\mathbf{e}$ is the vector of ones and $\cdot^{\diamond p}$ denotes a Hadamard power (the exponent applied element-wise). Recall from Section 5.2 that the covariance matrix $\mathbf{K}$ is required to be Positive Semi-Definite (PSD). We prove that as long as we ensure that all $x, x' \geq 0$, and $c > 0$, the rank 2 matrix $\mathbf{x}\mathbf{x}^\top + c\mathbf{e}\mathbf{e}^\top$ is PSD (we include the proof in the Supplementary Material for completeness).

It has further been shown that if a matrix $A \in \mathbb{R}^{n \times n}$ is PSD, then whether or not $A^{\diamond p}$ can be guaranteed to be PSD depends on the rank of A and the value of p [167]. Fitzgerald at al. [168] prove that if $\text{Rank}(A) \geq 2$, $A^{\circ p}$ is only PSD if $p \in \mathbf{N} \cup [n-2, \infty)$. Since in our case n is the total number of true demand observations (the size of $\mathcal{D}_T$), the inferred degree is very unlikely to be greater than or equal to $n-2$. We can therefore conclude that our proposed covariance function unfortunately does not, in general, result in a PSD covariance matrix $\mathbf{K}$. However, it is not uncommon for non-PSD kernels to be used nonetheless in applications where they perform well [165]. We therefore adopt the common strategy [169] of adding a sufficiently large perturbation to the spectrum of $\mathbf{K}$, such that its indefiniteness is no longer a problem. Though this 'artificial' shift causes bias in the resulting predictions, we do not find this to be an issue in practice.

In fact, there is an intuitive way of interpreting the bias introduced by this shift. From (5.6) in Section 5.2, and using the fact that $\nabla \log p(\mathbf{y}|\mathbf{f})|_{\mathbf{f}=\hat{\mathbf{f}}} = \mathbf{K}^{-1}\hat{\mathbf{f}}$ [157] we see that without a shift, our posterior predictive mean would be given by $\boldsymbol{\mu}_p = \mathbf{K}_*^\top \mathbf{K}^{-1}\hat{\mathbf{f}}$. When we shift the spectrum of $\mathbf{K}$ by adding some diagonal matrix $\mathbf{D}$, this becomes

$$\boldsymbol{\mu}_p = \mathbf{K}_*^\top (\mathbf{K} + \mathbf{D})^{-1}\hat{\mathbf{f}}. \tag{5.10}$$

Now let us compare this with the posterior predictive mean produced by an unshifted model, which instead uses a Gaussian likelihood. That is, assuming that observations include some additive noise which is normally distributed, $y_i \sim f(x_i) + \varepsilon$, where, $\varepsilon \sim \mathcal{N}(0, \sigma_n^2)$. In this case, the posterior predictive mean is

$$\boldsymbol{\mu}_p = \mathbf{K}_*^\top (\mathbf{K} + \sigma_n^2 I)^{-1}\mathbf{y},$$

which is very similar to the form of (5.10) with $\mathbf{y}$ having replaced $\hat{\mathbf{f}}$, and $\sigma_n^2 I$ replacing $\mathbf{D}$. In other words, we can understand the added shift $\mathbf{D}$ as adding an implicit assumption of a certain noise level in the data. In our case, we scale our inputs so that $x_i \in [0, 1]$, and use $\mathbf{D} = I$.

5.3.6 *Implementation*

To implement the GP regression method described above, we build upon the existing GPML MATLAB library created by Rasmussen and Williams [163]. The library is well-equipped with most GP functionality, and is modular, such that functions for the different components of GP regression are defined independently, making it possible to incorporate new features into the existing library. We extend GPML in two ways: first, we develop a new covariance function file to implement the variable degree polynomial covariance function defined in (5.9); second, we develop code to implement a quadrature method to approximate the integral given in (5.8), which is required to marginalise the hyperparameters. However, since this code is not currently vectorised, the computation time (especially for the quadratic approximation) is significantly longer than it could otherwise be (between 5 and 20 seconds depending on the fineness of the quadrature grid, with an AMD FX 4350 Quad-Core Processor).

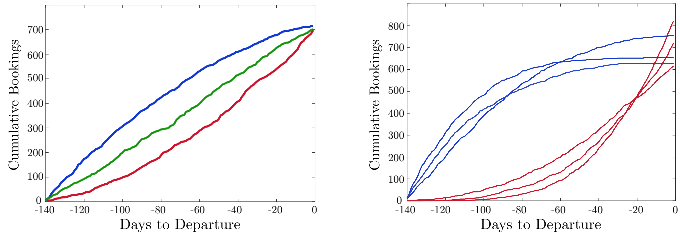
5.4 NUMERICAL EXPERIMENTS

5.4.1 *Experiment 1: 'The Queenan framework'*

We begin by reproducing the only experiment in the literature comparing the performance of the three best performing methods — DES, EM, and PD — presented by Queenan at al. [78]. They propose three typical cumulative demand curve types — convex, concave, and homogeneous (approximately linear), examples of which are shown in Figure 5.3a — and compare the performance of DES, EM, PD, LT, and N3 for the three curve types. In this section, we reproduce their experiment for DES, EM, PD (as well as the variants 'EM Daily' and 'PD Daily' described below) and compare these to our proposed GP regression unconstraining method.

5.4.1.1 *Constructing the test curves*

For each of the three curve types, 100 cumulative demand curves are generated, each with 140 data points, running from 140 days before departure up until the day before departure. To create one convex curve, daily demand for the first twenty days (140 – 121 days before departure) is sampled from a Poisson distribution with $\lambda = 2$. For the next twenty days, daily demand is sampled from a Poisson distribution with $\lambda = 3$, and so on, such that the final twenty days before departure have daily demand sampled from a Poisson distribution with $\lambda = 8$. This process is repeated 100 times to create the 100 convex curves. The creation of the 100 concave curves follows a similar procedure, the only difference being that the mean of the Poisson distribution begins instead at $\lambda = 8$ for the first 20 days, and is decremented by 1 every 20 days, such that demand over the 20 day period before departure is sampled from a Poisson distribution with $\lambda = 2$. In the homogeneous/linear case, daily demand is sampled from a Poisson distribution with $\lambda = 5$ for all 140 days before departure. We note that this process for generating test curves is equivalent to simulating a piecewise-homogeneous Poisson process, where the inter-arrival times within each day are exponentially distributed.



(a) Experiment 1 (b) Experiment 2

Figure 5.3: Sample demand curves generated in Experiments 1 and 2. Figure (a) shows a sample from the convex, concave and homogeneous curve sets (red, blue, and green, respectively) which are generated in Experiment 1. All 100 curves in each set have very similar shapes to the curves shown here. Figure (b) shows three samples from the convex and concave curve sets (red and blue, respectively) which are generated for Experiment 2. This illustrates the more realistic variation in the extent of convexity/concavity among curves used in Experiment 2.

The next step, for each curve type, is to calculate a set of 100 random booking limits, corresponding to each of the 100 curves. Those curves whose cumulative demand exceeds their corresponding generated booking limit are the constrained curves. In each case, booking limits are generated five times, in such a way that an increasing proportion of curves are constrained each time: the first set of booking limits constrains approximately 20% of the curves, the second set constrains approximately 40%, the third 60%, the fourth 80%, and finally the fifth set 98% of curves (see the Supplementary Material for details). These booking limits are used to artificially constrain the relevant booking curves. The various unconstraining methods are then applied to this artificially constrained data, producing unconstrained approximations, which are then compared with the 'true' generated data to evaluate their performance.

5.4.1.2 *Applying the unconstraining methods*

In order to apply the unconstraining methods, we first need to construct our set of demand data $\mathcal{D} = \mathcal{D}_T \cup \mathcal{D}_C$, using the curves and booking limits generated as described above.

For EM and PD this is done as follows: the set of true[1] cumulative demand totals for a given curve type is given by $\mathcal{A} = \{a_1, \ldots, a_{100}\}$, where a_i is the total cumulative demand of the i^{th} curve. For each set

1 Note that the term 'true' here and in Sections 5.4.2 and 5.4.3 refers to the fact that *in this experiment*, these are the demand values against which the uncontrained estimates will be compared. True demand in real-world scenarios may not be generated by Poisson processes as this experimental data is - see Section 5.4.3.

of booking limits $\mathcal{B}$, there is some subset $\mathcal{A}_C \subset \mathcal{A}$ containing those demand totals which are greater than (and therefore constrained by) their corresponding booking limits. We set $d_i = b_i$ for all constrained observations, and $d_i = a_i$ otherwise, which gives us our set $\mathcal{D}$, to which EM and PD[2] are applied (see the Supplementary Material for details). Note, therefore, that standard EM and PD are only applied to the cumulative bookings totals the day before departure.

It is also possible, however, to instead apply EM and PD to the daily bookings data from each day that any flight was constrained, thereby producing unconstrained approximations of the daily bookings for all constrained flights. Adding these approximations of bookings made on each constrained day to the bookings made prior to constraining, yields the total cumulative unconstrained approximation for that flight. We term this variation of the method "EM Daily" (and "PD Daily"). For EM and PD Daily, the procedure for assembling $\mathcal{D}$ is slightly different. The first step is to identify the first day that any curve in the set becomes constrained (that is, the earliest any curve exceeds its booking limit). We call this day $t_{\max}$. Next, a separate set $\mathcal{D} = \mathcal{D}_T \cup \mathcal{D}_C$ is created for each day t_k, starting from $t_{\max}$ up until departure. On each of these days, $\mathcal{D}_T$ contains the daily bookings from those curves whose cumulative totals by that day are still below their corresponding booking limit b_i. The set $\mathcal{D}_C$ contains all zeros (one for each curve which has surpassed its booking limit b_i by day t_k).

In the case of the single-curve methods, DES and our proposed GP regression method, each constrained curve is unconstrained independently. For each constrained curve, we calculate how many days before departure the booking limit was reached. In the case of DES, we then define $\mathcal{D}_T$ to contain the cumulative bookings up until that day, and all elements of $\mathcal{D}_C$ to be equal to the booking limit for that curve. Together these give us $\mathcal{D}$ and DES is applied to approximate $\mathcal{D}_U$ (see the Supplementary Material for details). For GP regression, we define $\mathcal{D}_T$ to contain the daily bookings up until the day of constraining, and all elements of $\mathcal{D}_C$ to be equal to zero. GP regression is applied in each case as described in Section 5.3. All four of DES, GP, EM Daily and PD Daily are therefore used to unconstrain data from the whole constrained period leading up to departure.

5.4.1.3 *Results*

Given the stochastic nature of the experiment, we repeat it five times and average the results. We present these results in Table 5.1, which mimics the format of those reported by Queenan et al. [78].

The success of each method is judged, as in [78], by calculating the difference (as a percentage) between (i) the mean of the set of

2 We apply PD using $\tau = 0.5$.

cumulative demand totals, where the unconstrained totals are used for those curves which were constrained, and (ii) the mean of $\mathcal{A}$, the true cumulative demand totals. We call this percentage mean error the E1 error.

The results in Table 5.1 show that when tested on the sets of convex and homogeneous curves, DES outperforms both EM and PD, though in the case of EM the margin is only significant when 98% of curves are constrained. This is unsurprising as it is well known that EM performs poorly when almost all data is constrained. It is also apparent that as the proportion of constrained data increases, the performance of PD deteriorates faster than that of EM. Once again, it is to be expected that these results diverge, because with less data points the conditional mean (used in EM) becomes less likely to be well approximated by the conditional median (used in PD).

	Proportion of Days Constrained				
	20%	40%	60%	80%	98%
			Convex		
EM	0.05	0.17	0.40	0.36	1.44
PD	0.22	0.54	0.77	1.36	2.79
DES	**0.03**	**0.07**	**0.06**	**0.06**	**0.21**
EM Daily	0.05	**0.07**	0.09	0.14	0.28
PD Daily	0.06	**0.07**	0.09	0.14	0.31
GP	0.06	0.20	0.23	0.31	0.42
			Concave		
EM	0.14	0.18	0.31	0.59	1.23
PD	0.29	0.48	0.74	1.55	3.52
DES	0.42	0.70	1.26	2.12	4.52
EM Daily	0.06	0.12	0.15	0.25	0.48
PD Daily	**0.03**	0.04	**0.08**	**0.08**	**0.31**
GP	0.06	**0.02**	0.12	0.30	0.52
			Homogeneous		
EM	0.11	0.41	0.23	0.20	1.40
PD	0.20	0.47	0.87	1.46	2.86
DES	**0.03**	0.07	0.04	**0.08**	0.12
EM Daily	0.04	**0.05**	0.08	0.17	0.29
PD Daily	0.04	0.06	0.10	0.18	0.38
GP	0.04	**0.05**	**0.03**	0.09	**0.07**

Table 5.1: Percentage mean error (E1) for Experiment 1. The numbers indicate the percentage error in the mean of the cumulative demand totals produced by each method (with the best performing in bold). GP regression is either comparable to, or better than, EM, PD, and DES, but according to this measure performs slightly worse in general than Daily EM and PD.

One noteworthy observation is that unlike all other methods, DES performed significantly worse on the concave set than on the convex set. The reason for this is that a given booking limit will constrain a concave curve for a longer period than a convex curve with the same total demand. Since DES only extrapolates linearly, it performs poorly on non-linear curves constrained for long periods of time.

The top three best performing methods across all curve shapes in this experiment are the daily variants of EM and PD, and our proposed GP regression method. All three produced very low percentage mean errors, which remained at or below 0.5% even when 98% of curves are constrained. In this experiment, GP is slightly outperformed by the other two. However, as we will see, this is more a result of unrealistic consistency across the curves in the test set than a reflection of the merits of daily EM and PD.

Assumptions behind the test curves

There are three important assumptions which underpin the creation of the curve test set described above:

1. The underlying 'arrivals' or bookings process for all curves can be modelled with either (i) a constant underlying Poisson rate λ or (ii) a linear $\lambda(t)$ approximated by a piecewise constant rate λ.
2. All curves of a given type are described by precisely the same (piecewise) constant Poisson rate λ.
3. The variance in the total demand among all curves will be quite small (for example, the mean total demand for the convex curves was 696.74 with a standard deviation of only 27.93).

A consequence of creating booking curves based on these assumptions that EM and PD perform as well as they do. Indeed, the resulting data set to which EM and PD are applied is by construction almost exactly normally distributed. Recall that EM and PD are based on the assumption that the underlying distribution of the data is normal, and it should therefore come as no surprise that they perform well when tested on a dataset which has, by construction, a distribution that is so close to normal.

However, the three assumptions listed above are often quite unrealistic. Firstly, our analysis of bookings data from Emirates Airlines shows that daily bookings are generally not well represented by a homogeneous Poisson process. Secondly, while modelling bookings as an inhomogeneous Poisson process is commonplace and sensible, the experiment allows for only (a crude approximation of) a linearly increasing or decreasing Poisson rate, and fails to model non-linearly changing underlying Poisson rates $\lambda(t)$.

Thirdly, it is extremely unlikely that an airline would be able to isolate (a priori) a set of historical booking curves for which the bookings follow exactly the same underlying inhomogeneous Poisson process. It is much more likely that even for flights with the same general demand shape (e.g. convex booking curves), the change in Poisson rate over time will vary — it might be linear for some flights, and perhaps approximately quadratic or cubic for others, for example.

Finally, even after restricting the set of flights under consideration to those that are believed to have similar booking trends historically, it is unlikely that the variance of total demand in this set would be so small (in the Emirates Airlines data set in Section 5.4.3, the standard deviation was approximately 20% of the mean total demand). The consequence of a test set with such a small variance is that EM continues to perform adequately even when curves are constrained for a long period of time, simply because it collapses into mean imputation (as shown in Section 5.4.2.2) and the mean is mostly within two standard deviations of each instance of total demand.

Given that these assumptions are unlikely to hold in realistic settings, the performance of EM and PD in this experiment should not be taken as evidence of their success in the airline industry.

Length of constrained period

Queenan at al. [78] follow Weatherford and Polt [82] in their decision to set the artificial curve-specific booking limits based on what proportion of the demand curves they wish to constrain. However, the method they use to set booking limits to achieve this goal ends up constraining most curves for only a few days before departure — even when 98% of convex curves were constrained, the average length for which they were constrained was approximately eight days.

Over periods as short as this, these cumulative demand curves can be fairly well approximated as linear, even though the overall trend might be decidedly non-linear. Since DES produces a linear extrapolation, it naturally performs well in this experiment on homogeneous and convex curves. However, in reality demand may well need to be unconstrained for a period over which the curve cannot be well approximated as linear. Such scenarios are tested by the concave case in this experiment, and unsurprisingly, in these cases DES performs much more poorly than the other methods. To properly evaluate DES, the method for setting booking limits should therefore focus on the length of the constrained period rather than exclusively on what proportion of demand curves are constrained.

Metric for comparing methods

To obtain the results presented in Table 5.1, we calculate the mean of the cumulative demand totals, where the unconstrained totals are used for those curves which were constrained, and compare it with the mean of $\mathcal{A}$, calculating the percentage error. However, given that one purpose of unconstraining is to predict what the true demand would have been for a particular flight whose demand was constrained, perhaps a more appropriate metric is to measure the difference between the total cumulative unconstrained value $\widehat{d_i}$ and the true value a_i, and consider the best method to be the one which minimises these absolute errors.

Crucially, the method rankings produced by using these two metrics are not necessarily the same. For example, when judged according to which method most accurately reproduced the true mean on convex curves with 98% of them constrained (E1 error, Table 5.1), DES outperforms GP with an E1 error of 0.21% as compared to 0.42%. However, if instead we consider the average absolute error in the unconstrained approximations of final cumulative demand (which we call the E3 error), GP performs similarly to DES (and even slightly outperforms it). The full set of E3 error results from Experiment 1 are shown in Table 1 in the Supplementary Material. Since accurately unconstraining each instance of constrained demand is a crucial function of unconstraining methods, considering only the percentage mean error compromises the assessment of the benefits of using each method in practice.

5.4.2 *Experiment 2: Generalised Queenan*

In this section we design and conduct a modified version of Experiment 1, creating sets of curves based on less restrictive assumptions, and use multiple metrics for adjudicating the relative performance of each method. We focus on the case of convex and concave curves, since truly homogeneous curves are much easier to unconstrain.

5.4.2.1 *Constructing the test curves*

In both the convex and concave case, we construct 90 demand curves, each with 140 data points (one for each day in the lead up to departure). The number of bookings d_t (i.e., on day t before departure) is sampled from a Poisson distribution with rate $\lambda(t)$. Instead of assuming the same underlying (piecewise constant) trend in λ for every curve, however, we model 30 curves as having a linearly changing $\lambda(t)$, 30 curves with a quadratic $\lambda(t)$, and the last 30 curves with a cubic $\lambda(t)$. The resulting set of test curves has a similar mean demand to those in Experiment 1, but is more realistic in two key ways: 1) there is larger variance in total cumulative demand among the curves (a standard deviation of approximately 65 (instead of 28) for convex curves, and 67 (instead of 27) for concave curves), and 2) there is more variation in the shape of the demand curves (see Figure 5.3b above).

We randomly select 15 out of each set of 30 curves to constrain, such that a total of 45 out of 90 curves of each shape are constrained. We repeat this process three times, constraining the curves for 5, 10 and then 20 days prior to departure. We use this process instead of the booking limits procedure followed by Queenan at al. because it enables us to control the length of the constrained period, ensuring it is sufficiently long, without constraining 100% of the curves (in which case we could not apply EM and PD). Our results show that it is not necessary to constrain as much as 80% or 98% of curves to illustrate the problems with EM and PD, and we restrict our attention to the case when only half the curves are constrained. As in Section 5.4.1, we repeat the whole experiment five times and average the results.

5.4.2.2 *Results*

Table 5.2 summarises the results of this second experiment. For the sake of completeness, we report three different error measures, each with their own motivation. These are as follows:

- **E1:** As before, we include the percentage mean error (E1), the error type reported in Queenan et al. [78] or Weatherford and Poelt [82], which measures the percentage error between the mean of the cumulative demand totals, where the unconstrained totals are used for those curves which were constrained, and the mean of the true cumulative demand totals.
- **E2:** We include the average absolute daily error (that is, the average distance between the true cumulative demand curve and the unconstrained curve during the constrained period), which we call E2. This indicates which method most accurately reproduces the constrained portion of the true demand curve, an important metric from the point of view of various airline applications (note that E2 cannot be measured for standard EM and PD because they are not applied to any data except the cumulative demand the day before departure).
- **E3:** We report the average absolute difference between the unconstrained approximation of the final cumulative demand and the true final cumulative demand of the constrained curves (E3).

These are shown in Table 5.2.

	Number of Days Constrained								
	5 Days			10 Days			20 Days		
Convex	E1	E2	E3	E1	E2	E3	E1	E2	E3
EM	2.05	-	51.8	0.59	-	58.16	**0.28**	-	57.55
PD	1.05	-	52.32	0.46	-	59.75	**0.28**	-	57.54
DES	0.50	7.50	11.47	1.32	12.56	23.52	4.12	28.42	63.77
EM Daily	0.15	13.76	22.62	0.19	22.23	41.60	0.36	33.98	70.23
PD Daily	**0.13**	13.78	22.63	**0.17**	22.23	41.56	0.33	34.00	70.18
GP	**0.13**	**5.87**	**8.13**	0.26	**8.70**	**14.29**	0.71	**16.38**	**31.43**
Concave	E1	E2	E3	E1	E2	E3	E1	E2	E3
EM	6.36	-	88.99	6.30	-	88.09	5.90	-	83.04
PD	4.67	-	65.33	4.67	-	65.34	4.34	-	61.12
DES	0.09	0.78	1.35	0.22	1.63	3.17	0.83	5.61	12.06
EM Daily	0.08	0.97	1.31	0.18	2.23	3.27	0.50	6.54	9.86
PD Daily	**0.05**	0.76	1.02	**0.13**	1.9	2.75	**0.30**	5.61	8.44
GP	**0.05**	**0.59**	**0.91**	0.14	**1.19**	**2.13**	0.31	**2.68**	**5.06**

Table 5.2: Results of Experiment 2. E1 denotes percentage mean error, E2 denotes average absolute daily error in the unconstrained approximation, and E3 denotes average absolute error in the total cumulative unconstrained approximation.

From Table 5.2 we see that GP regression outperforms every other method according to every measure, with four exceptions. The exceptions involve the E1 error measure on both curve sets: when the constraining period is 10 days long, E1 for GP is slightly (between 0.01% and 0.1%) larger than for the daily variants of EM and PD; and the same applies when 20 days were constrained, with the added outcome that E1 is also slightly lower for standard EM and PD than for GP on the convex curves. However, in both cases, E2 and E3 for GP is less than half what it is for all variants of EM and PD. We illustrate this stark contrast in performance in Figure 5.4a, which shows the probability density estimates of E3 error for EM, PD and GP when half the convex curve set was constrained for 20 days. These results highlight how misleading it is to consider only the percentage mean error (E1).

Moreover, the success of EM and its variants according to the E1 error measure on the convex curves is a result of the fact that EM collapses into mean imputation when the constrained cumulative demand curves are either very steep or constrained for a substantial period of time. To see why this happens, notice that if a demand curve is constrained for a significant period of time, the observed constrained demand is likely to be significantly lower than the mean of the true demand observations. The first step of EM calculates the initial estimates of unconstrained demand, $\widehat{d}_i^{(0)}$, as

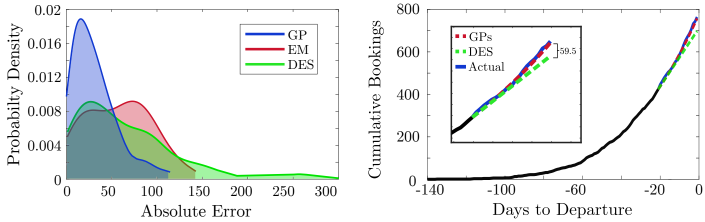
$$\widehat{d}_i^{(0)} = \mathbb{E}[\, d_i \,|\, d_i \geq b_i,\, d_i \sim \mathcal{N}(\mu^{(0)}, \sigma^{(0)}) \,], \text{ for all constrained } d_i, \tag{5.11}$$

$$= \frac{\int_{b_i}^{\infty} x\, p(d_i = x)\, \mathrm{d}x}{p(d_i \geq b_i)}, \text{ where } d_i \sim \mathcal{N}(\mu^{(0)}, \sigma^{(0)}). \tag{5.12}$$

As b_i decreases below $\mu^{(0)} - 2\sigma^{(0)}$, the probability in the denominator approaches 1, and the numerator approaches $\mathbb{E}[d_i] = \mu^{(0)}$. EM therefore predicts every constrained value as almost exactly $\mu^{(0)}$, and since $\mu^{(0)}$ is generally taken to be the mean of the true demand observations, this is equivalent to mean imputation [82].

For example, when EM is applied to the convex curve set when curves are constrained for 20 days, it produces unconstrained approximations for all 45 constrained curves which are extremely close to the mean of the true demand totals (all within 0.3 of this value). It is unsurprising that these estimates are very close to the mean of the true cumulative demand totals $\mathcal{A}$ given the way that the curves were constructed and how the subsets of curves to be constrained were selected (half of each curve shape). However, with the set of total demand having a standard deviation of 65, predicting the correct mean value for every unconstrained approximation results in a large absolute error on average (as shown in the E3 column of Table 5.2). The same explanation can be applied to explain the success of the daily version of EM in achieving a small E1 error, while resulting in a very large E2 and E3 error.

As expected, DES performs poorly on both convex and concave curves in this experiment, as both curve sets are constrained for the same length of time. Moreover, the increased curvature of some of the curves as compared with Experiment 1 exacerbates the failure of its linear extrapolation (see Figure 5.4b).

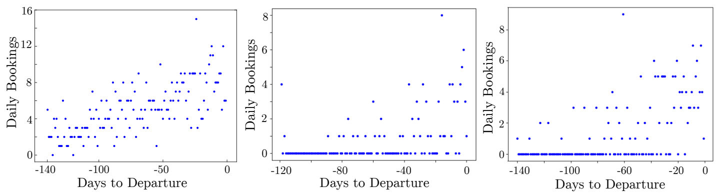


(a) E3 Error probability density estimates (b) Extrapolation using GP and DES

Figure 5.4: Figure (a) shows the probability density estimates of E3 error totals for the convex curve set in Experiment 2, with 20 days constrained. We include only the top performing methods, and exclude the results for PD as they coincide almost precisely with those of EM. Figure (b) shows a comparison of the unconstrained extrapolations produced by GP and DES on a convex curve from Experiment 2. The black line corresponds to cumulative bookings prior to constraining.

5.4.3 *Experiment 3: 'Double Poisson process' (DPP) data*

To design our final experiment, we analyse a data set of 392 demand curves for Emirates Airlines tickets for several given flight routes and for individual fare classes, with total cumulative demand above 70. In this section, we focus exclusively on demand curves of the convex type, as these are the most common in our data set.



(a) Experiment 1 (b) Real data (c) Experiment 3

Figure 5.5: Comparison of data generated for our experiments to real Emirates demand data. Figure (c), generated in Experiment 3 by the 'DPP' process, is notably more similar to the real data in Figure (b) than the data created by Queenan at al. which is shown in Figure (a).

Figure 5.5a shows a typical example of the demand data generated for Experiment 1. The data produced for Experiment 2 is similar, except for the fact that the average daily rate increases in either a linear, quadratic, or cubic manner, depending on the demand curve in question (see the Supplementary Material for figures). While the method used in Experiments 1 and 2 to model bookings as an inhomogeneous Poisson process is standard in the literature, the resulting test data it produces is, in fact, noticeably different to the typical daily demand data we see in our data set. This is immediately apparent when comparing Figure 5.5a with Figure 5.5b, which shows real data from a typical high demand flight.

This motivates an alternative approach to manufacturing test data which more closely replicates typical Emirates demand data. The benefits of such a method over simply comparing the methods on the real data itself, are two-fold:

1. It makes the experiment reproducible, such that it can be performed by any researcher, in particular those without access to proprietary Emirates Airlines demand data.
2. It avoids the problem of having to subset the real data into sets of demand curves which are considered to have similar demand trends (which is necessary prior to testing multi-curve methods, as they are only applied to such subsets). Airlines have their own ways of doing this subsetting, and the extent to which they are successful in creating appropriate subsets will greatly impact upon the performance of any multi-curve method. Manufacturing realistic data therefore allows us to control the bias which exists in favour of, or against, multi-curve methods while still producing industry-relevant results.

We started by attempting an alternative method for simulating an inhomogeneous Poisson process known as 'thinning'. However, we found that even our best efforts to generate realistic-looking data still fell short (see the Supplementary Material for details).

As an alternative starting point for designing more realistic test data, we note that one of the obvious differences between the data in Figure 5.5a and Figure 5.5b is that there are significantly more 0 data points (i.e., days on which 0 bookings were made) in the latter figure, and these continue until much closer to the day of departure. A first attempt on this basis might be to look at some of the literature on intermittent demand forecasting which accounts for frequent 0-bookings periods [170, 171, 172, 173] and try the methods used therein for simulating demand data. Unfortunately, the data produced with these methods do not display any trend, neither in inter-arrival times nor demand size, both of which are important features of the real airline data we have seen.

Instead, we propose an alternative data simulation method which does account for these trends, but one which still independently models two key features of the data:

1. The rate $\lambda(t)$ representing the trend in average daily bookings (on those days on which bookings are made); and

2. The 'inter-arrival times' (on the scale of days), that is, the number of days with zero bookings in between days on which bookings are made.

Modelling these two factors independently does indeed break the assumption of an underlying standard Poisson arrivals process. However, we find that doing so allows us to achieve more realistic-looking data nonetheless.

We propose sampling both 'arrivals' (bookings) and the 'inter-arrival times' (the number of days between bookings) from Poisson distributions, with time dependent rates $\lambda_1(t)$ and $\lambda_2(t)$ respectively.[3] To create a test curve, we therefore define two vectors $\boldsymbol{\lambda_1}, \boldsymbol{\lambda_2} \in \mathbb{R}_+^{140}$, with each element corresponding to the rate at each day in the lead up to departure. Thereafter we follow two steps: we use the vector $\boldsymbol{\lambda_2}$ to determine on which days non-zero bookings occur, and on each of these days we sample the number of bookings from a Poisson distribution with mean rate equal to the element of $\boldsymbol{\lambda_1}$ corresponding to that day (see Figure 3 in the Supplementary Material for an illustration of this process).

The data shown in Figure 5.5c was generated with this 'double Poisson process' (DPP), and it is evidently much more similar to those in the Emirates data set. We thus use this process to generate 90 convex demand curves of varying degrees of convexity as in Experiment 2 (the formulae for $\lambda_1(t)$ and $\lambda_2(t)$ used to create these curves are given in the Supplementary Material). The resulting curves have a mean total demand of 181.9 and a standard deviation of 35.2. We repeat this process three times, constraining 45 of the curves for 5, 10 and 20 days, and compare the performance of the various unconstraining methods in each case. As before, repeat this whole process five times, and we present the averaged results in Table 5.3.

3 We note that the Poisson distribution is an especially unusual choice for modelling inter-arrival times, since it means that they must be integer valued. However, since we are operating on the scale of 'days', and producing a data set which gives daily bookings, this property of the Poisson distribution does not cause any problems: we hope to model precisely the integer number of days between days on which bookings occur.

	Number of Days Constrained								
	5 Days			10 Days			20 Days		
	E1	E2	E3	E1	E2	E3	E1	E2	E3
EM	7.06	-	30.04	3.22	-	26.45	1.10	-	29.08
PD	4.51	-	26.88	1.48	-	26.57	0.86	-	29.20
DES	0.98	4.02	6.14	2.64	6.93	12.30	8.15	13.88	30.81
EM Daily	0.59	5.32	8.65	1.15	7.67	13.69	2.30	12.27	24.58
PD Daily	**0.31**	5.05	8.12	0.44	7.22	12.78	**0.49**	10.94	22.36
GP	**0.31**	**3.39**	**4.88**	**0.40**	**5.60**	**9.19**	1.46	**9.56**	**18.65**

Table 5.3: Results of Experiment 3. As in Table 5.2, E1 denotes percentage mean error, E2 denotes average absolute daily error in the unconstrained approximation, and E3 denotes average absolute error in the total cumulative unconstrained approximation.

Once again, GP regression outperforms all other methods, in each case and by almost every measure. The only exception is that we see a repeat of the result that we saw in Experiment 2, where the E1 error for a constrained period of 20 days is slightly higher for GP than it is for EM, PD and PD Daily. However, GP is again noticeably better according to both the E2 and E3 error measures, which we argue are also relevant for evaluating the performance of the various methods, for reasons discussed in Section 5.4.1.

5.4.4 *Experimental conclusion*

The results from all three experiments described in this section indicate that our proposed GP unconstraining method achieves significantly better results on average than existing state of the art methods in the literature.

5.5 DETECTING CHANGEPOINTS IN DEMAND TRENDS

Thus far, we have assumed that the underlying trend in demand is likely to be generally smooth, without kinks and discontinuities, and we have compared the performance of GPs to existing unconstraining methods using test data which stays true to this assumption. In this section, we relax this smoothness assumption, allowing for points at which the characteristics of the underlying demand trend change dramatically.

In time-series analysis, times at which the characteristics of the data change dramatically are known as *changepoints*. Research on methods for detecting changepoints has been ongoing for decades [174, 175], and a range of methods have been applied to data from a diverse set of subjects ranging from hydrology [176] to the history of political relationships [177].

Demand for flight tickets is affected by many exogenous factors, some of which (like a competitor's prices or the perceived safety of a destination) can change drastically in short spaces of time. This makes it likely that changepoints will feature frequently in the demand data, and we have observed them in the Emirates data discussed in Section 5.4.3. Any method which hopes to accurately pick up and forecast a demand trend based on past time-series data will make significant errors, unless it is able to account for the possibility of changepoints. Without registering a changepoint, the method's extrapolation will be informed by all past data, when in fact it needs to account for the fact that the data from *before* the changepoint does not accurately represent the demand trend *after* it.

Garnett at al. [161] showed how changepoint detection can be simply and elegantly incorporated into a GP regression framework by constructing an appropriate covariance function. For our purposes, we want to allow for the fact that the covariance before and after the changepoint might be completely different (for example, a roughly quadratically increasing demand level before the changepoint, and a small, approximately constant demand level after the changepoint). We therefore define our covariance function to be

$$k(x, x') = \begin{cases} \sigma_1^2 (x^\top x' + c_1)^{p_1} & \text{if } x, x' < x_c, \\ \sigma_2^2 (x^\top x' + c_2)^{p_2} & \text{if } x, x' \geq x_c, \\ 0 & \text{otherwise,} \end{cases} \tag{5.13}$$

where x_c is the location of the change point, and is also a hyperparameter which is inferred from the data along with the other covariance function hyperparameters $\theta_c = \{\sigma_1, \sigma_2, c_1, c_2, p_1, p_2, x_c\}$. We implement this by adding a further covariance function file to the GPML library, and we marginalise out all hyperparameters using the quadrature method presented at the end of Section 5.2.

We omit comparisons with other methods in our analysis of changepoint detection. Since detecting changepoints is exclusively applicable when demand data is viewed as time-series data, direct comparisons with multi-curve methods like EM and PD do not make sense. Moreover, the fundamental problem with DES (that it is limited to linear extrapolation of cumulative demand) would apply equally in the presence of changepoints. Instead, we illustrate the power of our method with a number of scenario-inspired test cases, with normalised inputs $x \in [0, 1]$, the results of which are shown in Figure 5.6. In particular, we consider the demand for a given fare-class on a given flight on Airline A in the following three scenarios:

- Scenario 1 (Figures 5.6a and 5.6b): Demand is fairly constant but relatively low because the price of the ticket is not competitive. When a competitors' price suddenly increases or their flights are sold out, this diverts demand to Airline A, causing a jump in the average daily bookings.

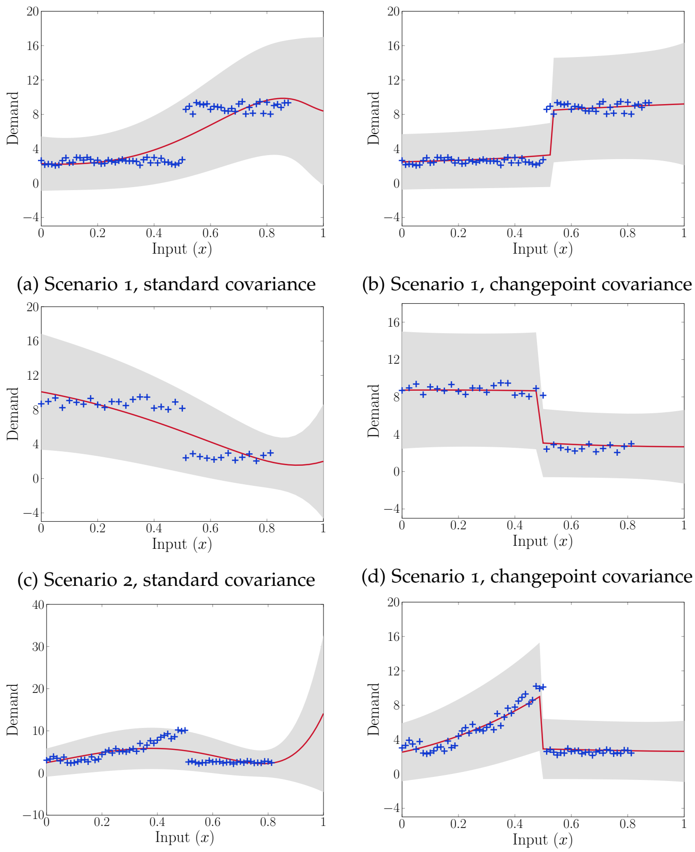
- Scenario 2 (Figures 5.6c and 5.6d): The inverse of Scenario 1, where Airline A's flight begins with high relatively constant demand, until a cheaper option on a competitor airline becomes available, diverting demand away from Airline A.

- Scenario 3 (Figures 5.6e and 5.6f): Demand for Airline A's flight is increasing in some non-linear fashion until there is some destination-related shock (a headline news item causing the destination to be considered unsafe, for example), after which demand drops dramatically, and remains fairly flat at a relatively low level.

The plots on the left-hand side (Figures 5.6a, 5.6c, and 5.6e) show the results of fitting our proposed GP regression model with the standard covariance function given in (5.9), which does not account for changepoints. We compare these with the plots shown on the right-hand side (Figures 5.6b, 5.6d, and 5.6f) which are obtained using the changepoint covariance function defined in (5.13) on the same data. It is clear in all cases that the changepoint covariance function better extrapolates the post-changepoint demand trend, and Figure 5.6e best exhibits just how wrong the GP prediction might become when a changepoint is not accounted for. In this case, demand has clearly collapsed to a low level, and there is no indication that it is likely to pick up again. The changepoint covariance function picks this up and correctly extrapolates this low, relatively constant demand. When the changepoint is not accounted for however, the predicted future demand begins to increase rapidly in strong contrast to the post-changepoint trend.

For illustration purposes, we have thus far only included examples of data with a single changepoint.[4] However, our approach can easily be extended to cope with multiple changepoints given sufficient computational power (Garnett at al. [161] show that in fact, when taking a 'moving window' approach to changepoint detection, it is uncommon to need to account for more than two changepoints in the covariance function itself). This ability to detect and account for discontinuities

4 Though the existence of a changepoint is pre-specified when defining the changepoint covariance function (5.13), if it transpires that there is in fact no changepoint, the inferred covariance parameters on either side of x_c will be roughly the same, resulting in a prediction similar to that made by the standard, non-changepoint covariance function given in (5.9)



(a) Scenario 1, standard covariance

(b) Scenario 1, changepoint covariance

(c) Scenario 2, standard covariance

(d) Scenario 1, changepoint covariance

(e) Scenario 3, standard covariance

(f) Scenario 3, changepoint covariance

Figure 5.6: Prediction and extrapolation in the presence of changepoints. In each scenario, the data is the same in the figures on the left and right. The figures on the left show the GP fit and extrapolation using the covariance function given in (5.9), while the figures on the right use the modified changepoint covariance function defined in (5.13).

in the demand trend is a powerful motivating factor for the use of our method for unconstraining demand.

5.6 DISCUSSION

In this chapter, we proposed and extensively tested a new single-class unconstraining method that uses GP regression with a Poisson likelihood and a new covariance function — a 'variable degree polynomial covariance function'. This proposed regression model is novel in its inference of the (non-integer) exponent in the polynomial covariance function, and its ability to perform well on more realistic demand data; data which exhibits nonlinear inhomogeneous Poisson rates, varying inter-arrival-times and discontinuities in demand.

The results of our numerical experiments point to a number of important conclusions. The first is that it is important to consider not only the percentage mean error (as is standard in the literature), but also the average absolute error in the unconstrained approximations. For some applications of unconstraining; say, in estimating average historical demand for pricing and route-planning purposes, accurately estimating the mean of historical demand may well be the most important goal. However, the decision about whether or not to re-open a currently-constrained fare class depends on accurately unconstraining that particular demand curve, not knowledge of the historical mean, and hence what matters in such cases is the accuracy of individual unconstrained approximations. Importantly, our results show that success according to one measure does not imply success according to the other, and hence in general we ought to consider both. Secondly, our results show that when both error measures are considered, our proposed GP method notably outperforms all other unconstraining methods included in our experiments. Our GP method performs comparably (or better) when considering the percentage mean error, and performs decidedly the best when considering average absolute error. One notable drawback was that with un-vectorised code, the use of GPs is significantly more computationally expensive than other methods. However, we expect that with vectorised and optimised code, GPs would be much more computationally comparable.

Our final piece of analysis extends our proposed method to cope with changepoints in time-series demand data, by defining an amended covariance function, and inferring both the location of the changepoint and the covariance hyperparameters on either side of it from the data. We show that extrapolation based on data which exhibits a changepoint is much more accurate when the amended (changepoint) covariance function is used.

While these results are very positive, they also highlight areas which merit further research. First, it would be preferable (though non-trivial) to find a set of conditions under which we can guarantee that the variable degree polynomial covariance function we propose will work with a predetermined spectral shift. Secondly, though we find GPs to be preferable to all other methods tested in our experiments, the superior performance of EM Daily and PD Daily in Experiments 1 and 3, as compared with the standard average-based EM and PD, merits further investigation by RM practitioners into increasing the granularity of the data on which they apply these methods.

5.7 CONCLUSION

In summary, we have the following observations and conclusions:

1. Most state of the art single-class unconstraining methods make restrictive assumptions on the distribution of demand.
2. Rather than relying on historical observations of demand in DCPs, we propose a GP regression method that extrapolates the booking curve.
3. Especially toward the end of the booking curve, when the probability of class closure is high (and therefore, limited data points), this method outperforms traditional methods.
4. We show that even when a booking curve is constrained up to 98% over time, our model produces reliable unconstrained booking curves.
5. We find that this method is robust against non-linear demand trends, variations in total demand, lengthy periods of constraining, non-exponential inter-arrival times, and discontinuities and changepoints in demand data.

6 ITINERARY-BASED BOOKING PREDICTION

Abstract

Demand forecasting is extremely important in RM. After all, it is one of the inputs to an optimisation method which aim is to maximize revenue. Most, if not all, forecasting methods use historical data to forecast the future, disregarding the "why". In this chapter, we combine data from multiple sources, including competitor data, pricing, social media, safety and airline reviews. Next, we study five competitor pricing movements that, we hypothesize, affect customer behavior when presented a set of itineraries. Using real airline data for ten different OD-pairs and by means of Extreme Gradient Boosting, we show that customer behavior can be categorized into price-sensitive, schedule-sensitive and comfort ODs. Through a simulation study, we show that this model produces forecasts that result in higher revenue than traditional, time series forecasts.

This chapter is based on [20].

6.1 INTRODUCTION AND MOTIVATION

Our definition of RM, introduced in Section 2.1, is the *process of dynamically assigning capacity to products of perishable nature with a fixed total capacity*. In practice, this means determining what booking classes should be open, for what origin and destination (OD) pair such that the overall, network revenue is maximized. For this optimisation process, we need the airline's own demand forecast, fares and capacity.

Traditionally, it was thought it was sufficient to segment a market by whether a customer is a business or leisure passenger. An example of this can be found in fare rules: since the inception of pricing, the "Saturday rule" has been used. This rule says that a customer has to stay at least a Saturday night before returning to their origin. Business travellers want to spend their weekends at home, while leisure passengers do not mind spending a Saturday night.

Teichert et al. [46] however, show that this type of segmentation is not sufficient. They find customers that travel in business class for non-business reasons, and customers in economy class travelling for business reasons. Instead, they define five different segments: efficiency, comfort, price, price/performance and all-round performance.

Moreover, they find that it is difficult to segment customers, but rather *trips* should be segmented. For example, someone that travels for business may not be price sensitive, but when this same individual travels for leisure, they are. As the industry has changed, whether it is through deregulation, or through the advances in data capture and analytics, segmentation on the other hand, has not.

Modelling customer behavior is complicated, and there are many reasons for this. First, everyone is different: everyone prioritizes aspects differently. Second, not everyone acts rational. While it is impossible to model these characteristics, it is important to gain an understanding of underlying processes. This can help us understand why people make certain decisions.

In this chapter, we investigate the effects of competition on booking behavior in order to make itinerary-based booking predictions. We do so by combining several data sources. An overview of these data sources is given in the next section. Using these data sources, we engage in feature engineering. Next, we divide these features into those that are airline specific (for example, safety record), while others are itinerary specific (for example, departure time). The objective of this chapter is to build a model that given a set of itineraries, to predict what itineraries will be purchased. Armed with this data, airlines can then use this as a strategic tool to increase their demand.

We consider reviews of a well-known airline review website. This dataset consists of the actual review text, as well as ratings given by the user to the seat, in-flight entertainment (IFE), meal, crew and ground service. Next, we analyze the last 10000 tweets of the airlines that appear in this article. Based on these two sources, we perform sentiment analysis.

Another vital dataset gave us access to (historical) pricing information. For a given OD, this dataset captures the price for every airline for every departure date at every day before departure (DBD). Visually, these curves not only tell us what the price was at what point in time before departure, we can also inspect whether airlines react to each other's price change.

We were given access to a data source that includes information about the airline, such as fleet size, fleet age and total aircraft worth. We also have access to an airline safety index.

Lastly, we have OD-specific characteristics. These are features engineered from the OTA's search results. Features include whether this OD has a day flight, whether there is a direct flight, the time of the first departure of the day, the time of the last departure of the day, the number of frequencies, the minimum connection time and the minimum travel time (= flying time + connection time).

This chapter is organized as follows. We provide an overview of the data used for our work in Section 6.2. Our approach to this problem is discussed in Section 6.3, before modelling is covered in Section 6.4. We review the model's performance in Section 6.5. A discussion of our work and directions for further research are given in Section 6.6. In Section 6.7 we provide conclusions. In the Appendix, in Section 6.8, we show all engineered features, its source, and calculation.

6.2 DATA OVERVIEW

In this section, we provide a brief overview of the data sources we used.

6.2.1 *Online Travel Agent (OTA) dataset*

This data source contains both search queries and bookings made by customers. Customers on this website are tracked through cookies, as well as through their accounts (if they are logged in). In this chapter, we only look into actual bookings made, so we omit details of search queries. A sample of the dataset is shown below.

od	airline_id	dep_day_id	dbd	dep_time_mam	travel_time	price
AMS-LHR	4	3063	-119	1305	4.33	173.92
AMS-LHR	4	3213	-3	465	2.86	225.46
AMS-LHR	4	3444	-101	870	7.96	178.81
AMS-LHR	1	3448	-83	420	6.82	228.74
AMS-LHR	2	3481	-33	805	0.9	363.20
AMS-LHR	3	3621	-40	1265	0.9	425.37
AMS-LHR	4	3625	-98	420	5.81	132.23
AMS-LHR	3	3677	-100	835	0.9	453.85
AMS-LHR	3	3966	-47	865	0.9	140.14
AMS-LHR	2	3966	-47	440	0.9	277.91

Table 6.1: Sample of booking dataset

As we see from Table 6.1, we are given itinerary details of bookings. For storage purposes, this company does not store competitor offerings that were not booked.

While the true dataset contains the actual airline name, departure date and other revealing details, this OTA asked to obfuscate airline names and dates for this study, as they consider it sensitive information.

6.2.2 *Competitive pricing*

The data supplied here contains the historical price for every itinerary offered by every airline. This data is similar in terms of dimensions of the data given in Section 6.2.1, but contain pricing information of *all* itineraries, not just the one that were purchased. An example of this dataset is given below, in Table 6.2.

od	airline_id	dep_day_id	dbd	dep_time_mam	travel_time	price
FRA-SYD	1	946	-6	1220	13.24	605.73
FRA-SYD	2	946	-6	1200	15.83	416.74
FRA-SYD	3	946	-6	445	12.95	336.32
FRA-SYD	4	946	-6	455	12.65	719.43
FRA-SYD	5	946	-6	800	13.72	634.05
FRA-SYD	6	946	-6	815	10.5	795.12
FRA-SYD	7	946	-6	445	15.41	564.63
FRA-SYD	8	946	-6	1290	14.99	677.94
FRA-SYD	9	946	-6	800	14.75	582.23

Table 6.2: Example of competitive pricing data

Note that the data in Table 6.2 greatly enriches the data from the previous section - for every booking made, we now know how each competitor's schedule and price compared. Naturally, the fare in this dataset should, in theory, match with the fare that is associated with the booking. For the vast majority of the bookings (93% of our dataset is within 1%) this is the case : we refer to the reader to Figure 6.1. We review this in our discussion section.

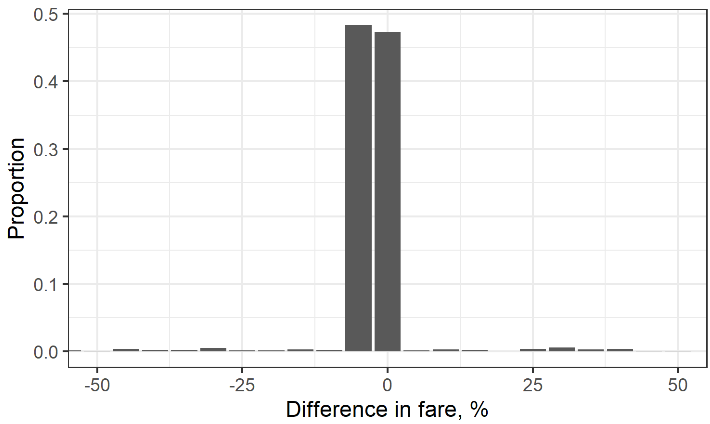


Figure 6.1: Fare errors in our dataset

We used the same encoding for obfuscating the airline names and dates. To make comparing fares fair, we always use the fares from this dataset (even though it is given for bookings in our other dataset).

6.2.3 *Airline ratings*

The website we used for our data contains reviews on airlines. People are given the opportunity to write a (free text) review, as well as rate their trip based on several characteristics, shown below.

id	airline	rec.	review	fb	ground	ife	crew	seat	value	wifi
1	5	N	(..)	2	3	3	2	1	4	1
2	1	N	(..)	2	2	2	1	1	3	2
3	2	Y	(..)	3	4	4	3	4	4	5
4	4	Y	(..)	5	4	5	5	5	5	4
5	5	Y	(..)	2	2	3	5	4	4	4

Table 6.3: Airline review dataset. We have omitted the review text to save space. Rec is short for recommended.

Passengers rate their airline based on a general recommendation, their F&B offering, the service on the ground, the in-flight entertainment (IFE), the quality of the seat, value for money and WiFi performance. They also have the possibility to write a free-text review.

6.2.4 *Twitter sentiment*

We were given access to the past year's worth of tweets of the airlines present in our data. This dataset required a lot of preparing: we left out retweets (RTs), replies, and we only focused on reviews written in English. Retweets were left out to avoid duplicates records and create a bias towards sentiment. Replies, most often by the airline, are not representative of an individual's perception and for this reason these were left out. Finally, we only focuses on reviews written in English since these would not require translation: translating tweets from other languages may lose the impact they had in their native language.

6.2.5 *Airline safety ratings*

This website provides an index for airline ratings. It uses accident and incident history, environmental factors and operational risk factors to derive a safety score. A sample of this data is given in Table 6.4.

rank	airline_code	score
1	CX	0.005
2	NZ	0.007
3	HU	0.009
4	QR	0.009
5	KL	0.011

Table 6.4: Airline safety index

6.2.6 *Fleet information*

This dataset contains information on the airline's fleet. It contains several properties of an airline's fleet, such as size and cost. While not always correct, we aim to use the average fleet's age as a proxy for a comfort rating (newer aircraft are typically quieter and provide better entertainment). Similarly, we intend to use the fleet size as a proxy to how well passengers are accommodated when irregular operations happen (if an airline only has a handful of aircraft and a flight gets cancelled, it is likely a passenger will endure long delays). An example of this data is shown in Table 6.5.

airline_id	aircraft	aircraft cost	aircraft registration	aircraft age
1	77W	300	PH-ABC	8.8
1	77W	300	PH-XYZ	12.2

Table 6.5: Fleet information example

In Table 6.5, we show the airline ID, again, obfuscated to anonymize the data, the IATA aircraft code (for example, 77W represents a Boeing 777-300ER), its obfuscated registration, and aircraft age in years.

6.2.7 *Data overview*

In this section, we provide a few characteristics of our dataset. Table 6.6 shows the number of competitors by OD. This is not an exhaustive list of all airlines that sell this OD. Rather, it is a list of airlines that sold this OD in our dataset.

AMS-DXB	AMS-LHR	AMS-SYD	CDG-SYD	FRA-SYD	FRA-KUL	FRA-SYD	KUL-SIN	LHR-JFK	LHR-SYD
7	4	5	4	9	6	5	2	2	5

Table 6.6: Number of competitors by OD

Figure 6.2 shows the demand (sum of bookings) of competitors by OD. Note not every airline is operating every OD. For example, it is unlikely that an airline operating KUL-SIN will also operate LHR-JFK.

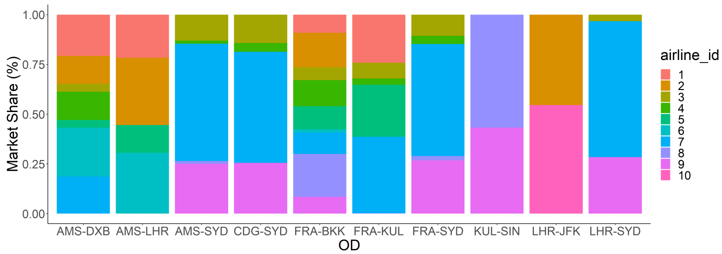


Figure 6.2: Demand by OD and competitor

Figure 6.3 illustrates the average travel time in hours by airline by OD.

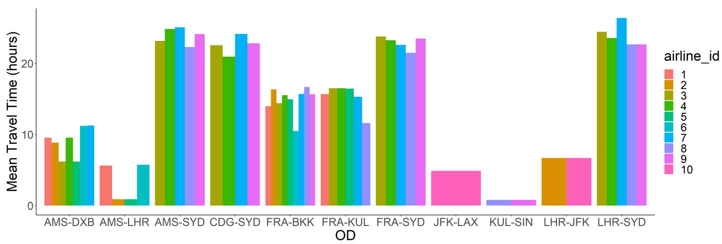


Figure 6.3: Travel time by OD and competitor

Taking a look at AMS-LHR, the big difference can be explained by the fact that airline 2 and 3 offer a direct flight between these airports, while airlines 1 and 4 offer connections, resulting in longer average travel times. Note that for LHR-SYD the small difference in travel times across airlines. Even though there are five competing airlines in this OD, each with their own hub, and a thus a different way to fly, the resulting travel time is quite similar.

6.3 APPROACH

In this section, we will illustrate the approach we have taken. First, we reiterate the objective in Section 6.3.1. Next, in Section 6.3.2 we discuss the features we have engineered.

6.3.1 *Objective*

Consider an itinerary, with features as described in the previous section. The objective is to determine whether an itinerary will be purchased. We will use Extreme Gradient Boosting to build this binary logistic model (after all, an itinerary is purchased yes or no). Note that this is different from typical choice modelling. In that case, the objective is given a set of alternatives, what option a customer is most likely to purchase. Our objective goes beyond that - we look at the features of an individual itinerary to determine whether this will be purchased.

Since we suspect that behavior will be different across ODs, we suspect we cannot compare a short-haul to a long-haul OD, or a business-heavy to a leisure-heavy OD, we build a separate model for every OD.

6.3.2 *Feature engineering*

Arguably, the most important step in building the model is the process of feature engineering. Feature engineering "involves constructing novel features from given data (...) driven by domain knowledge developed over time (...) with the goal of improving predictive learning performance." [178]. In this section, we follow the same structure as in the Data Overview section, Section 6.2, and will review our feature engineering.

6.3.2.1 *OTA dataset*

From this dataset, we construct the following **airline/OD-specific features**:

1. Number of unique itineraries offered
2. Minimum, maximum and average flying time
3. Minimum, maximum and average travel time (flying time+connection time)
4. Is an evening departure offered (boolean, if there is an itinerary departure time after 6PM)
5. Departure time of first flight of the day (first sector), in minutes after midnight
6. Departure time of last flight of the day (first sector), in minutes after midnight
7. Arrival time of first flight of the day (first sector), in minutes after midnight

8. Arrival time of last flight of the day (first sector), in minutes after midnight
9. Is a direct flight offered? (boolean, true if and only if there is an itinerary with one sector)

Next, we derive the following **itinerary-specific features**:

1. Flying time, in hours
2. Travel time, in hours
3. Morning, afternoon or evening departure time: departure time for the first sector before 9AM, before 6PM, or after 6PM respectively
4. Morning, afternoon or evening arrival time: arrival time for the first sector before 9AM, before 6PM, or after 6PM respectively
5. Wide-body or narrow-body aircraft used
6. Night flight (boolean, if departure time is before midnight and arrival time is after midnight, both in local timezones).

6.3.2.2 *Competitive pricing*

Using this dataset, we derived features that measure current and rolling window price movements. In the example that follows, we look into the calculation of pricing features for airline 1 for 100 days before departure. The features compare this airline's pricing against the cheapest airline in the market, dubbed yy, and second cheapest airline in the market, dubbed xx. We use this terminology, since airline codes typically consist of two characters.

Table 6.7 shows a subset of fares for four different airlines at specific times before departure.

time	airline_1 fare	airline_2 fare	airline_3 fare	airline_4 fare
-103	450	600	500	1000
-102	475	600	500	1100
-101	450	625	360	1100
-100	450	750	500	300

Table 6.7: Example of fares

time	airline_1 fare	yy fare	xx fare	is_cheapest?	yy fare difference	xx fare difference
-103	450	450 (1)	500 (3)	Y	0	-50
-102	475	475 (1)	500 (3)	Y	0	-25
-101	450	360 (3)	450 (1)	N	90	0
-100	450	300 (4)	450 (1)	N	100	0

Table 6.8: Step 1 - Calculation of fare difference. The number between parentheses indicates what airline this fare belongs to.

Table 6.8 shows how we first calculate the yy and xx fare. This is simply the cheapest and second cheapest fare. Note that the parentheses indicate what airline offers that specific fare. Our first engineered feature is *is_cheapest*, which is a Boolean and is true if the airline's fare equals the yy fare. Note that you can have multiple airlines that are cheapest, if they have the same fare. Next, we calculate the difference between the airline and the yy and xx fares.

time	mean3d_yy	mean_3d_xx	farediff
-100	(0+0+90)/3=30	(-50-25+0)/3=-25	450-360=90

Table 6.9: Step 2 - Feature engineering

Table 6.9 then illustrates how these features are calculated. Note that the $mean3d_{yy}$ and $mean3d_{xx}$ fares indicate the average difference in fares over the past three days, while $farediff$ is a snapshot feature that only measures the difference at only one particular day ($t = -99$).

Apart from calculating the mean, we also calculate the standard deviation. We repeat this process for a 7, 14 and 28 rolling window. As a result, we will have the following features:

1. $mean3d_{yy}, sd3d_{yy}, mean3d_{xx}, sd3d_{xx}$
2. $mean7d_{yy}, sd7d_{yy}, mean7d_{xx}, sd7d_{xx}$
3. $mean14d_{yy}, sd14d_{yy}, mean14d_{xx}, sd14d_{xx}$
4. $mean28d_{yy}, sd28d_{yy}, mean28d_{xx}, sd28d_{xx}$
5. *farediff*.
6. *is_cheapest*

6.3.2.3 *Airline ratings*

From our dataset, we derived median values for the characteristics, as well as counted the observations. These are given in Table 6.10.

Some reviews contain what route the reviewer flew. Ideally, one would only look into reviews that match the OD we are studying. However, the resulting number of reviews are too low to be reliable

airline	rec.	review	f&b	ground	ife	crew	seat	value	wifi	obs
1	0.89	6.52	4.19	4.18	3.6	4.65	3.8	4.1	2.6	79
2	0.52	5.84	3.22	3.04	2.79	3.64	3.25	3.25	2.66	638
3	0.74	6.21	3.62	3.77	4.44	3.73	3.97	3.75	3.63	221
4	0.6	5.63	3.15	2.94	3.29	3.37	3.31	3.06	3.2	87
5	0.89	6.35	3.84	3.89	3.35	4.19	3.86	4.2	2.67	71

Table 6.10: Example of airline ratings

measures for aggregates. For this reason, we have chosen to take aggregates across airlines. Note in Table 6.10 that there is no OD present. We review this decision in our discussion.

Table 6.10 shows whether passengers recommend this airline, as well as the median scores for the onboard F&B, the service provided on the ground, the IFE, the crew, quality of the seat, value for money and WiFi. The "obs" column show how many reviews we collected.

The value under review is constructed using sentiment analysis. We do this as follows. The free text of all reviews are read into R. Each review is converted into a long $1 * N$ vector by splitting the review; each element in the vector will have a single word. First, stop words and punctuation are removed. The text is converted to lowercase. This vector may then be joined with the AFINN dataset. The AFINN dataset [179] was created by Nielsen containing 1468 unique words, that Nielsen manually labeled with a score between minus five (highly negative) and plus five (highly positive). Next, we simply take the mean over this list of scores to determine how positive or negative a review was, and scale this value between 0 and 10. An example of the AFINN dataset is shown in Table 6.11.

word	score
amazing	4
breathtaking	5
disaster	-2
distrust	-3
excellence	3
fraudsters	-4
limited	-1
misleading	-3

Table 6.11: AFINN subset sample

6.3.2.4 *Twitter sentiment*

After removing quotes, retweets, special characters among other things, we followed the same procedure as for airline ratings reviews: we matched every word in the tweets with the AFINN list, then took an average of these scores to get a rating by tweet. We aggregated these individual reviews by taking the median of each tweet's rating to obtain a score by airline, then scaled it to a value between 0 and 10. Just like for airline ratings, we were unable to obtain scores by OD - we were unable to derive from the tweets what OD passengers were flying. An example of this dataset is given in Table 6.12.

airline_id	twitter_sentiment
1	6.32
2	4.22
3	7.02
4	6.49
5	7.21

Table 6.12: Twitter sentiment scores by airline

6.3.2.5 *Airline safety Ratings*

We did not engineer any features. Instead, we used the score provided to us.

6.3.2.6 *Fleet information*

For fleet information, we derived the following **airline-specific features**:

1. Fleet size: number of aircraft
2. Fleet cost: sum of aircraft cost (list price)
3. Fleet age: median of aircraft age

An example of these features is shown in 6.13.

airline_id	fleet size	fleet cost	fleet median age
a	226	42343	11.19
b	268	54130	11.67
c	249	85298	6.15

Table 6.13: Fleet information example

In Table 6.13 we have used a different $airline_{id}$ as in previous sections, as there are ways to trace the actual airline's name using these characteristics.

6.4 MODELLING

Having engineered features, we will predict whether a given itinerary is purchased. To accomplish this, we will use extreme gradient boosting.

6.4.1 *Extreme gradient boosting*

In this section, we provide a brief overview of extreme gradient boosting (XGB). For a full introduction of XGB we refer the reader to Chen et al [180]. In what follows, we provide a short, alternative brief. Suppose we have an input x_i, and an output y_i. We would like to make a prediction, denote this by $\widehat{y}_i$.

An example of estimating y_i is given in Equation (6.1).

$$\widehat{y}_i = \sum_{j=1} \alpha_j x_{ij} \tag{6.1}$$

This, of course, is simple linear regression (we omit a constant and standard error): in this case y_i is expressed as a linear combination of explanatory variables, denoted by x_i. The objective is to estimate those α_j that minimize an error measure. We typically want to minimize some error measure, denoted by O_m, depending on a model m. A natural selection of an objective function is an error measure:

$$O_m(\alpha) = \sqrt{\sum_{j=1} (y_i(\alpha) - \widehat{y}_i)^2}. \tag{6.2}$$

This is known as the root mean squared error (RMSE).

Extreme Gradient Boosting is a tree boosting algorithm. The method works in a similar fashion: first, specify how our predictor is expressed in terms of features, like (6.1). Next, specify an objective function, comparable to (6.2). Finally, iterate to find the optimal value.

Consider the following objective function:

$$O_{XGB}(\theta) = L(\theta) + \Omega(\theta) \tag{6.3}$$

In Equation (6.3), the first term, $L(\theta)$ is the loss function. Typically, (R)MSE is used. The second term, $\Omega(\theta)$, is the regularization term.

This term measures the complexity of the model and helps us control overfitting.

Let $f_k(x_i)$ be a function that takes a set of input parameters x_i, as before, and return the score of x_i in tree k, $k = 1,..,K$. Suppose we have M different features. In XGB, we assume that

$$\widehat{y_i} = \sum_{k=1}^{K} f_k(x_i). \tag{6.4}$$

That is, the prediction for y_i, $\widehat{y_i}$, is the sum of linear combination of the score in each tree. Let T be the number of leaves in a tree, and w_i the weight of leaf i. Assume that we have M features. Therefore, our input x_i is a M-dimensional vector. Introduce a function $q(x_i)$ which takes an input x_i, and follow the rules of the decision tree to map it to the leaves. Specifically:

$$q(x_i), \mathbb{R}^M \longrightarrow T$$

The prediction is then given by this function $q(x_i)$, weighted by the weights of the leaves, denoted by w_i. Therefore:

$$f(x_i) = q(x_i)w_i$$

Since XGB is an iterative method, there is no static objective function as in Equation (6.2). Similarly, the prediction of y_i at time t is given by the previous value of y, represented as y_i^t, plus the score of x_i in our new tree:

$$\widehat{y_i^t} = \widehat{y_i^{t-1}} + f_t(x_i) \tag{6.5}$$

Suppose now we have a generic loss function, l, some choice for L, as we introduced in (6.3). In this case, we have:

$$O_{xgb}^t = \sum_{i=1}^{n} \left(l(y_i, \widehat{y_i^{t-1}} + f_t(x_i)) \right) + \Omega(f_k) \tag{6.6}$$

XGB uses a second-order Taylor expansion to approximate this function l. Recall that the Taylor expansion of $f(x)$ at $x + a$ up to the second degree is given by:

$$f(x) = f(x+a) + \frac{f'(x+a)}{1!}(x+a) + \frac{f''(x+a)}{2!}(x+a)^2 \tag{6.7}$$

Deciding to use the MSE for our generic function l, our objective function O_{XGB} at time t is equal to:

$$O_{XGB}^t = \sum_{i=1}^{n} \left(y_i - (\widehat{y_i^{t-1}} + f_t(x_i)) \right)^2 + \Omega(f_k) \tag{6.8}$$

$$= \sum_{i=1}^{n} \left(2(\widehat{y_i^{t-1}} - y_i) f_t(x_i) + f_t(x_i)^2 \right) + \Omega(f_k). \tag{6.9}$$

The term that remains is $\Omega(f_k)$. As we discussed above, this term is important but often forgotten, and helps us control the complexity of the models by penalizing large models. In XGB, this function is defined as follows:

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum_{i=1}^{n}(w_i)^2. \tag{6.10}$$

In the regularization shown in (6.10), γ is threshold of reduction in the loss function for XGB to further split a leaf. Smaller values will make XGB split more leaves, therefore generating a more complex tree structure, while larger values will limit the number of leaves. We chose $\gamma = 0.25$. On the other hand, λ penalizes on large values of w_i. Intuitively, this is an appealing property: it encourages XGB to use all of its inputs a little bit, rather that some of its inputs a lot. The choice of λ is defined by the user. In our work, we chose $\lambda = 1$. Other parameters are investigated in Section 6.4.2.

6.4.2 *Selecting parameters*

Learning Rate, η
The learning rate is the shrinkage used. The shrinkage factor is a way to slow down the incremental performance gain of a new tree being added to the ensemble. A smaller learning rate means the model will take longer to run but is less likely to overfit. We try values of $\eta = 0.01, 0.02, .., 0.1$.

Number of Decision Trees, n_t
The number of decision trees specifies how many trees can be used until we stop optimizing. In practice, this number is typically relatively low, 1000 or less (depending on the model size) and is a direct result of how the algorithm works. More specifically, it is the result of how fast errors are being corrected. A new boosted tree model is constructed based on errors of the current tree. We therefore expect to see diminishing returns. Let n_t be the number of trees we can use. We perform a grid search over values of $n_t = 50, 100, .., 500$.

Depth of the tree, d_t
This parameter specifies how many layers a tree may have. Intuitively, a small number of layers in a tree do not capture enough details about the data to be a good descriptor. On the other hand, a tree with too many levels may be overfitting the dataset. Let d_t be the depth of the tree. We will evaluate values of $d_t = 3, 4, ..., 20$.

Subsample, s_t
Subsample represents the percentage of the number of observations chosen to fit a tree. Taking out too much data means the model will run faster (after all, less fitting needs to be done), but taking not enough data may expose us to overfitting. Let s_t be the proportion of data used to fit a tree. Then we will try $s_t = 0.2, 0.3, ..., 1$.

Number of features used per tree, f_t
In the dataset, each row represent an observation. Every column contains a feature. The XGB algorithm samples the number of columns in building a new tree. Using all columns for every tree may lead to overfitting, but also makes the problem slower to solve.

We present the results in Section 6.5. More specifically, in Section 6.5.4, we will review the influence of each of these parameters for the different ODs.

6.4.3 *Performance*

Suppose we have the following, generalized linear model that is used to make a prediction for a value y_i:

$$\hat{y}_i = \sum_j w_j x_{ij} \tag{6.11}$$

In our dataset, we have an exhaustive list of options offered by the most popular airlines on this route. We know what itinerary was purchased. This will be our label. As a result, we have a logistic binary objective function: yes or no.

The extreme gradient boosting algorithm will return a probability of purchase – between 0 and 1. To obtain a yes/no label, we round the probability to the nearest integer. We used the R implementation of the widely used *xgboost* package, specifically version 0.72. We used a maximum tree depth of 20, and used 10 passes (iterations) of the data.

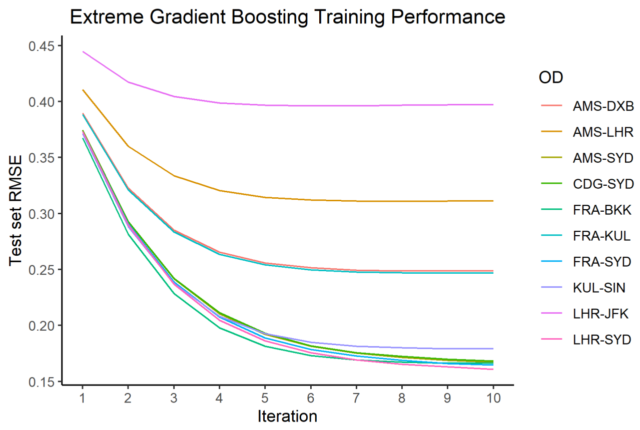


Figure 6.4: RMSE error improvement by iteration

Figure 6.4 shows the RMSE by iteration. Note from this graph that the RMSE decreases quickly. We show that the number of passes chosen, 10 is sufficient to obtain sufficiently close to optimal performance, while making sure runtime is acceptable.

6.4.4 *Runtime*

The XGB model performance is excellent in terms of runtime - we record an average of 11.9 seconds across the 10 ODs we consider.

On the other hand, the feature engineering process takes a significantly longer amount of time. This involves a series of expensive operations on the dataset. Calculating the pricing features, in particular, involves joining a dataset on itself. This was done in memory in R. Despite this, the entire process, from loading data to the actual engineering of features took an average of 23 minutes per OD. This is important metric, since frequent forecasting and reoptimisation (discussed in 2) is important in practice.

6.5 RESULTS

6.5.1 *Comparison with logistic regression*

In this section, we compare the performance of the proposed model against the logistic regression model. The results are shown in Table 6.14. Note that the model is biased against predicting a non-purchase, since the majority of itineraries are never purchased. We therefore omit true negatives from our performance metric, and study the proportion of false negative, false positive, and true positive predictions.

Classification	False Negative		False Positive		True Positive	
OD / Method	Logit	XGB	Logit	XGB	Logit	XGB
AMS-DXB	0.55	**0.34**	**0.05**	0.19	0.39	**0.47**
AMS-LHR	0.40	**0.26**	**0.10**	0.18	0.50	**0.56**
AMS-SYD	0.63	**0.11**	**0.05**	0.06	0.32	**0.83**
CDG-SYD	0.39	**0.09**	0.10	**0.06**	0.51	**0.86**
FRA-SYD	0.24	**0.17**	**0.07**	0.13	0.69	**0.70**
FRA-KUL	0.36	**0.23**	**0.11**	0.19	0.53	**0.58**
FRA-SYD	0.80	**0.10**	**0.04**	0.06	0.16	**0.83**
KUL-SIN	0.09	**0.07**	**0.00**	0.02	0.90	**0.91**
LHR-JFK	0.39	**0.20**	**0.10**	0.17	0.51	**0.62**
LHR-SYD	0.63	**0.10**	**0.05**	0.06	0.32	**0.84**

Table 6.14: Comparison of False Negatives, False Positive and True Positives for the logistic regression (logit) and XGB model. Bolded values represent the better value

Table 6.14 shows the performance by OD. The numbers in bold face compare the logit and XGB model by prediction type and highlight the better value. First, let us consider the false negative: the XGB model outperforms the logit model for every OD. The differences range from 0.02 for KUL-SIN to 0.7 for FRA-SYD. We suspect that the logit model for the KUL-SIN OD performs relatively well since the number of competitors is low (namely, 2) and seems to be driven by a single feature, which we will discuss in Section 6.5.3.

Reviewing the false positives, there is only one OD in which the number of false positives are lower for the XGB model. The differences range from outperforming logit by 4% to 14% more false positive predictions for AMS-DXB. Comparing these false positives to false negatives, the results seem to indicate that the logitistic regression model is biased toward predicting false negatives, while the XGB model is biased toward predicting false positives.

However, most interesting are the true positives. In all cases, the XGB model outperforms the logit model. This seems to illustrate the need for a more advanced method than simple logistic regression. The differences in performance range from 1% on the KUL-SIN OD, to 67% on FRA-SYD. For FRA-SYD, note that the number of false negatives is almost equivalent to the number of true positives. In fact, the performance gains on most SYD ODs are impressive: 51% for AMS-SYD, 15% for CDG-SYD and 52% for LHR-SYD.

6.5.2 *Overall performance*

As before, it should be noted that only 20.3% of all options displayed were purchased. For this reason, to study the effectiveness of our model, we disregard true negatives. The percentages, in what follows, are calculated by comparing the element against true positives, false negatives and false positives. This is shown in Figure 6.5.

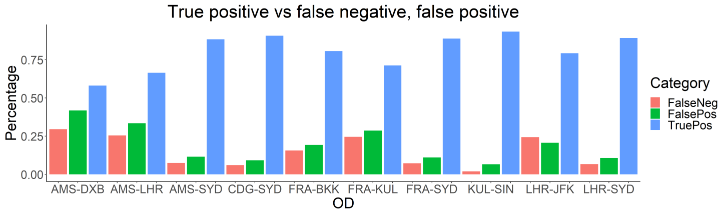


Figure 6.5: Model performance by OD

Note the difference in performance between ODs. For example, for AMS-DXB, we note a relatively high rate of false positives and false negatives. Our initial hypothesis was this is caused by the number of competitors: the more options presented to the customer, the more dispersed the data, and therefore the more mistakes we can make. However, from Table 6.6 we note that this is not the case. For example, consider the AMS-DXB with AMS-LHR ODs. These have respectively seven and four competitors. However, the number of false negatives and false positives in terms of percentage for AMS-LHR is barely different. Now compare LHR-JFK with KUL-SIN. Both of these ODs have two competitors, but the model performs much better for the latter OD. We have therefore evidence which may suggest that the LHR-JFK market is one that has more dynamics than the KUL-SIN market.

Overall, the results across ODs are very positive. This is illustrated by the median percentage of 84.5%. The best performing OD is KUL-SIN with a score of 93.4%, closely followed by the SYD ODs. AMS-LHR and AMS-DXB with 66.5% and 58.1% respectively score worst.

Complete confusion matrices are show in the appendix. Out of all ODs, we feel that the AMS-DXB case is the most worrying. Comparing 1769 true positives with 1275 false negatives, the model severely underestimates the total number of bookings. These ratios are much lower for other ODs.

6.5.3 *Customer behavior*

In our opinion, the more interesting topic is what features drive booking behavior. We refer the reader to Figure 6.6. This figure shows, by OD, the gain of each feature. We exclude any features with a gain smaller than 0.05.

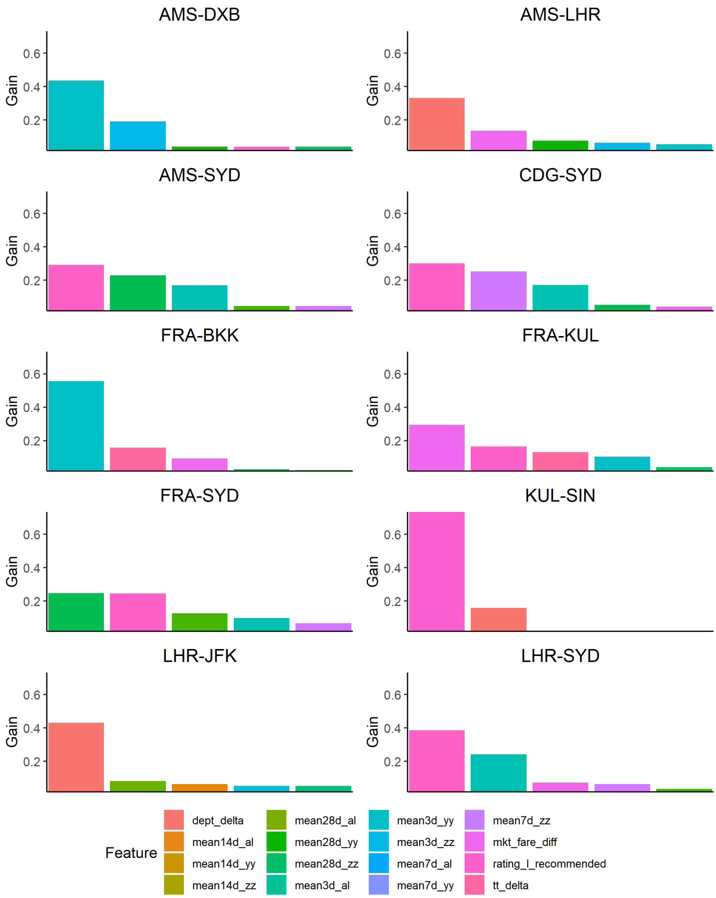


Figure 6.6: Gain of features by OD, sorted by importance from left to right

We note three very different behaviors: price sensitive ODs (OD pairs that are predominantly affected by price), departure-time sensitive ODs (those OD pairs that are driven by schedule) and comfort ODs (OD pairs for which passenger comfort is important). These are discussed in Sections 6.5.3.1, 6.5.3.2 and 6.5.3.3 respectively.

6.5.3.1 *Price sensitive ODs*

The pricing features dominate booking behavior. Interestingly, for AMS-DXB, the features of how the airline moves against the cheapest and second cheapest airline in the market is dominant; the actual price difference to that cheapest airline is not. This seems to illustrate the fact that customers are set on booking with a particular airline, but monitor how this preferred airline is pricing itself against those cheaper airlines. Another example of this is FRA-SYD. The feature that captures the price movement against the cheapest airline in the market is dominant. On the other hand, the FRA-KUL market seems to be concerned with the fare different against the cheapest airline. The same applies for KUL-SIN – it only seems to be driven by price. To summarize, the OD's in this category are AMS-DXB, FRA-SYD, FRA-KUL and KUL-SIN.

6.5.3.2 *Departure time sensitive ODs*

Amsterdam, London and New York are traditionally considered routes for business travellers. One of the reasons for this is that all of these cities are financial hubs. It is worth noting that the delta in time between the airline's departure time and 6AM has a lot of explanatory power. Note that especially in the LHR-JFK case, the pricing features have little weight. In the AMS-LHR case, the fare difference versus the cheapest airline in the market is the second most powerful variable. We hypothesize that this is as a result of the recent increase in low cost carrier (LCC) frequencies between Amsterdam and London. In summary, the OD's that make up this category are AMS-LHR and LHR-JFK.

6.5.3.3 *Flying comfort ODs*

On these ultra long-haul ODs, a passenger's comfort is a deciding factor. Note that the review scores of the IFE have great explanatory power, and actually have greater explanatory power than the price. Also note that other features that may describe the comfort of a journey, such as the quality of the seat, crew or ground services did not appear. This seems to indicate that on long-haul ODs, passengers value their entertainment more than their seat! While airlines traditionally segment their pricing based on the origin, it is worth nothing that for these ODs terminating in Sydney customer behavior seems fairly consistent. Summarizing, the OD's that fall in this category are: AMS-SYD, CDG-SYD, FRA-SYD and LHR-SYD.

6.5.3.4 *Comparing ODs*

It is interesting to compare the model's performance, shown in Figure 6.5 with the features used to obtain the predictions, shown in Figure 6.6.

Consider the AMS-SYD and CDG-SYD ODs. From Table 6.6, we note that these ODs have five and four competitors, respectively. Looking at Figure 6.2, we note that the distribution of demand by airlines is similar: Airline 8 is missing from CDG-SYD. Airline 4 has proportionally more demand for CDG-SYD because of this missing airline. In short, from a competition perspective, we can argue that these are similar. Looking at the model's performance in Figure 6.5, we note very similar results for all three metrics: false positive, false negative and true positive. The metrics that powered these predictions, in Figure 6.6, show that the information gain for both the rating of IFE and $mean3d_yy$ features are very similar and these combined have an information gain of 0.45, with IFE being the most important feature. For this reason, we have segmented these together as a "comfort" OD.

6.5.4 *XGB performance*

In this section, we will review the parameters of the XGB model we introduced in Section 6.4.2. Note that the number of rounds is not shown here, but shown in Figure 6.4 and discussed in Section 6.4.3.

Learning Rate, η
Figure 6.7 shows the different values of learning rate η for different ODs. Recall that this parameter, also known as shrinkage, controls how much weight a new tree is assigned.

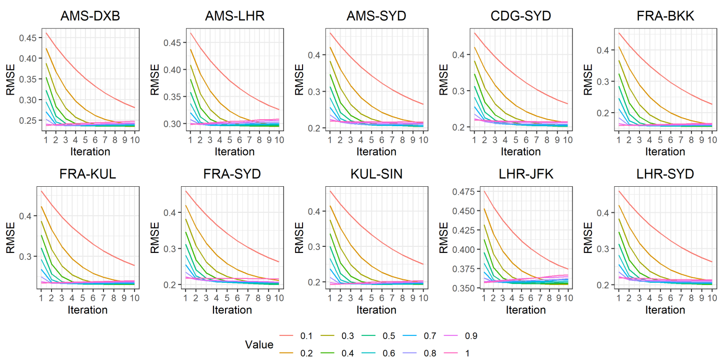


Figure 6.7: Performance of learning rate η by iteration and OD

From Figure 6.7, as expected, we see a very slow improvement of RMSE for $\eta = 0.1$ (red line). Increasing this parameter to $\eta = 0.2$ (orange) shows a much faster convergence. Values of around $\eta = 0.3$ (moss green) seem to be the best trade off between finding a good RMSE and runtime. For this reason, $\eta = 0.3$ is chosen. Note that very large choices of η actually results in a slight increase of RMSE as the number of iterations grow across the different ODs. In conclusion, we do not observe great differences between the different ODs, and keep the parameter fixed at $\eta = 0.3$ for all ODs.

Depth of the tree, d_t

Figure 6.8 shows the development of RMSE on the test set for different values of the depth of the tree we allow, d_t.

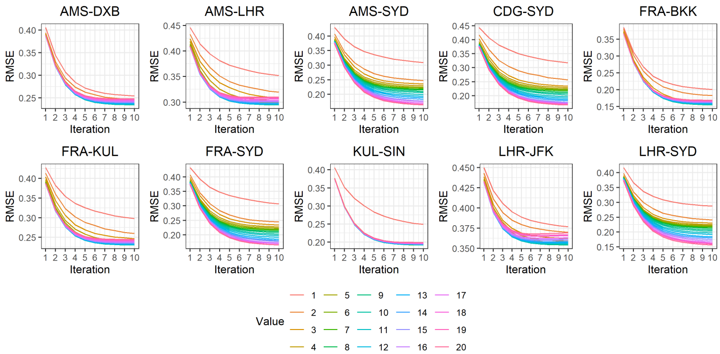


Figure 6.8: Performance of d_t by iteration and OD

First, let us consider the KUL-SIN OD. Here, we see a dramatic improvement from $d_t = 1$ (red), a tree with one level, to $d_t = 2$ (dark orange), a tree with two. The performance of this model, measured as the RMSE over the test set, no longer improves as we grow d_t further. This seems to indicate that the ensemble of trees for this OD is relatively simple. Comparing the different ODs, we may conclude that different ODs have different levels of complexities: for the AMS-SYD, CDG-SYD, FRA-SYD and LHR-SYD ODs, we see a clear improvement as we grow d_t up to its maximum (chosen) value of $d_t = 20$, while other ODs seem to converge at values of d_t beyond 7 (light blue). The value of d_t depends on the OD and is chosen visually, at the lowest value after which we do not see any improvement of performance. For example, for AMS-SYD we choose $d_t = 5$ (moss green), since we do not see any improvement for larger values of d_t, while for LHR-SYD, we choose $d_t = 20$ (bright pink).

Number of observations, s_t

In Figure 6.9, we study the effects of the number of observations used when building trees.

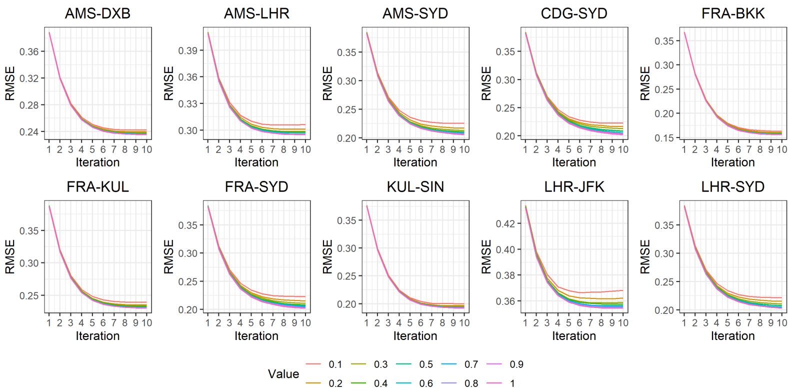


Figure 6.9: Performance of s_t by iteration and OD

From Figure 6.9, we observe that the parameter s_t is important for some ODs, while it is not important for others. For example, consider AMS-DXB. Earlier, we saw that the value of d_t was important for this OD, while it does not seem to be of importance for s_t, even for very small values of s_t. This may indicate that the observations for this OD are fairly uniform: the model performs just as well randomly selecting 10% of all observations as it does using all observations. Other ODs, such as FRA-SYD, seem to exhibit different model performance for different values. We choose the smallest value of s_t accordingly: this helps reducing both runtime and avoids overfitting. For AMS-DXB, we choose $s_t = 0.2$ (orange), and for ODs that are affected by the number of observations, such as LHR-JFK, we choose $s_t = 0.1$ (red).

Number of features, f_t

Figure 6.10 shows the effect of RMSE over ten iterations when using a certain proportion f_t of features when constructing trees for different ODs.

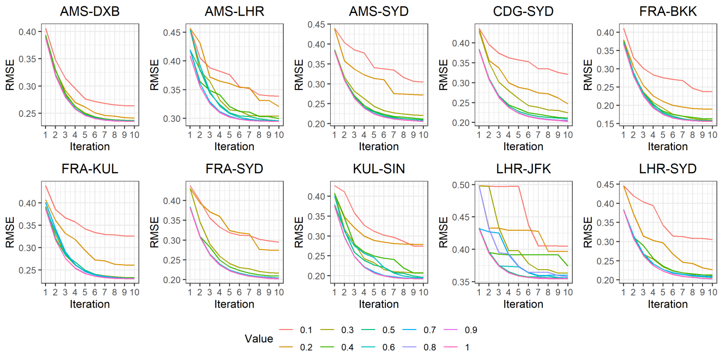


Figure 6.10: Performance of f_t by iteration and OD

In practice, a value of $f_t = 0.1$ (red) means that we use less than two features when modelling. This exemplifies the poor performance in terms of RMSE for all ODs. In Figure 6.6 we observed that for most ODs, two features are often very important in the performance of the model. Not (randomly) choosing these two features will therefore naturally result in poor performance. While the performance of the model does improve for low values of f_t, the trade-off is not worth it: even when choosing $f_t = 1$, the longest runtime across the 10 different ODs is less than four seconds. In summary, we choose $f_t = 1$ for all ODs.

6.5.5 *Examples*

One of the main drawbacks of XGB is that method is a *black box* method. This method produces an ensemble of decision trees. While a single decision tree is easy to understand, an ensemble is not. In this section, we use the *xgboostExplainer* [181] package to show, for a number of examples, how the model arrives at its prediction. This is achieved by drawing the ensemble of all trees, and traversing them to obtain the probability estimation.

We will review three examples below. Each of these examples represents a given itinerary. The probability of purchase is shown on the vertical axis. The numbers inside the bar represent the log odds of each feature. As discussed in the approach section, our goal is to predict whether a itinerary is purchased. For this reason, we round the probability to the nearest integer. As a result, the horizontal line at $p = 0.5$ represents the cut-off for predicting whether or not this itinerary is purchased or not. Features are ordered by their weight from left to right.

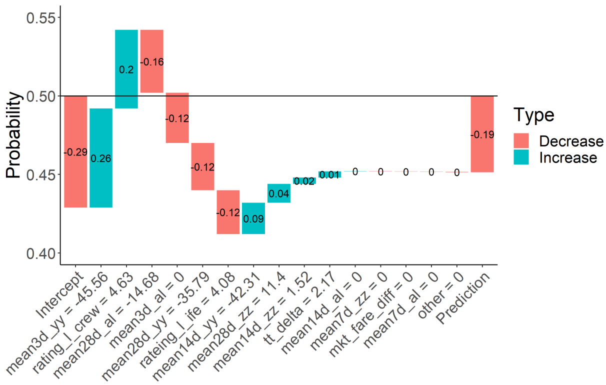


Figure 6.11: Example one: long-haul itinerary, low-priced airline

Figure 6.11 shows en example of how the probability is generated for a long-haul itinerary, for a low-priced airline. The model starts at a probability of purchase of 43%, the intercept. This probability increases by 6% to 49%, because this airline was $46 cheaper than the cheapest airline in this OD market. The crew for this airline is rated very high, at 4.63/5, which causes the probability of purchase to strengthen to 55%. This probability drops back to just over 50% because the airline itself has been cheaper in the past 28 days, on average by $14, indicating a price sensitive market. Note that the IFE score of 4.08 drags the purchase probability down by almost 3%. The final probability of purchase is 0.45, which, after rounding, is marked as a non-purchase.

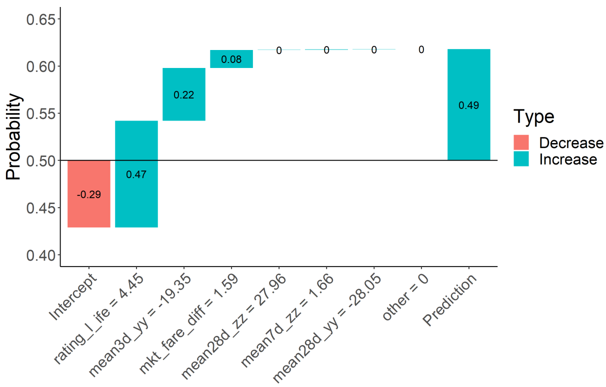


Figure 6.12: Example two: long-haul itinerary, well-priced airline with great IFE

Figure 6.12 shows another example. In this example, the airline is well-priced compared to competition and possesses a great IFE product. The baseline purchase probability is again 43%. Here we clearly see the impact of IFE, the probability of purchase increases by over 11%. Compare this to the impact the pricing difference to the cheapest carrier in the market in the past three days: the log odd impact of the IFE rating weighs twice as heavy as this pricing feature (0.47 and 0.22, respectively). Other features have negligible impact. The model generates a purchase probability of over 62%.

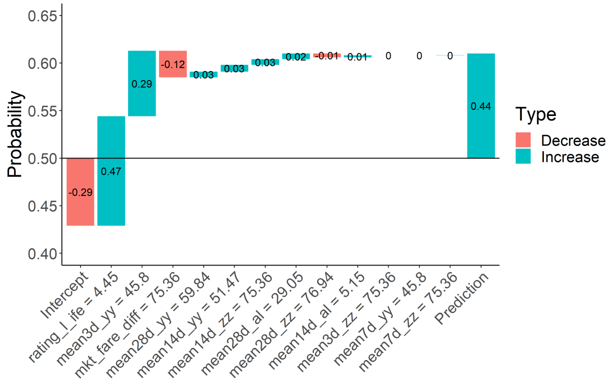


Figure 6.13: Example three: long-haul itinerary, higher-priced airline with great IFE

Figure 6.13 shows another example of how the model generates its purchase probability. In this case, the airline has great IFE but is higher priced than competitors. The baseline purchase probability is 43%. Again, we see how the IFE has the biggest impact on determining purchase probability. Interestingly, the airline is on average by $45 more expensive than the cheapest carrier in the market, this actually *increases* the purchase probability. This could indicate the willingness to pay more for a itinerary that contains a better IFE product. However, it also shows a bound - note that because this carrier is $75 more expensive than the market fare at time of purchase, the purchase probability declines slightly 2%. This may indicate that the sweet spot this airline could demand lies anywhere between $45 and $75.

6.5.6 *In practice*

As we discussed, predicting whether an itinerary is purchased can provide great insights in the positioning of the airline. It can benefit pricing and marketing teams and help them understand how to each of the features affect a customer's decision.

However, to understand the true value of our work, consider what the end objective of RM is: maximizing revenue. Studies, for example, [2], have shown that increases in forecast accuracy show clear increases in revenue performance. Intuitively, this makes sense: inputs closer to reality should mean better output revenues.

Therefore, in this section we will compare demand forecasting through our framework with demand forecasting typically done in practice. In practice, this is done by forecasting aggregated OD/class (for brevity) demand. To compare revenues, we will use a time series forecast in simulations to provide us with a base case. Next, we use the framework in this chapter to predict demand on an itinerary-level, then aggregate this to OD/class-level and use this as input in simulations for a single flight.

Unfortunately, we were only given access to data for 10 ODs. In practice, however, a flight has many different ODs crossing it. We are unable to create forecasts for every OD. Therefore, we have chosen a flight from a big airline that has four out of the ten ODs utilizing it. These four ODs combined represent, on average, 42% of the total demand for this flight.

The simulation is constructed as follows:

1. Calculate an optimal policy using a traditional forecasting method, in our case we chose double exponential smoothing;
2. Calculate an optimal policy using our new forecasting framework discussed in this chapter for the ODs that are available, use the traditional forecasting method for those ODs we do not have data for;
3. Generate $n = 500$ arrival processes, with an average demand factor of 0.98 and standard deviation of 0.1;
4. Simulate accepting requests and calculate revenues for both methods.

In the simulations, we make a distinction between whether a customer downsells or not. Downsell occurs when a customer purchases a cheaper class than she was willing to pay for. Suppose we have twelve different classes for sale, with ordered fares such that f_1 is the most expensive product and fare f_{12} the cheapest. Now consider, for example, a customer wanting to purchase booking class 10 with fare f_{10}. However, class 12 is available for sale with fare f_{12}. A customer downsells if she actually purchases class 12. In this case, the airline loses $f_{10} - f_{12}$. We have done extensive research into downsell and have covered this in Chapter 8 and in the literature [22].

The fares for the four ODs we have used are covered in Table 6.15. The associated demand by fare brand is given in Table 6.16. Demand by booking class has a relatively high variance, so we have summarized these in terms of fare brand. Recall that the definition of a fare brand was introduced in Chapter 2: fare brands are a collection of products with identical fare conditions, with the only differing factor price.

OD \Class	1	2	3	4	5	6	7	8	9	10	11	12
1	2324	1913	1672	1152	1081	966	871	706	660	498	494	447
2	2489	2078	1995	1707	1462	1363	1187	1009	774	553	534	474
3	1904	1621	1323	1094	1091	962	922	737	682	622	495	311
4	2509	2043	1452	1420	1035	762	700	523	449	374	311	206

Table 6.15: Fare by booking class

OD \Fare Brand	1	2	3
1	0.05	0.30	0.65
2	0.10	0.18	0.70
3	0.02	0.46	0.40
4	0.12	0.20	0.59

Table 6.16: Percentage of demand by fare brand. Booking Classes 1 through 3 are part of Fare Brand 1, 4 through 8 part of Fare Brand 2, 9 through 12 part of Fare Brand 3.

Table 6.17 shows the simulation results.

Downsell	Std	XGB	% Gain
No	27468	27648	0.70
Yes	9622	9898	2.90

Table 6.17: Revenue performance comparing forecast methods Std, Standard forecasting, with XGB. Demand factor = 0.98

In Table 6.17, we compare the standard forecasting technique (Std) with XGB. The numbers represent the revenue performance for this flight. The percentage gain shows the relative performance. If customers do not downsell, our simulations show an average improvement of 0.7%. However, of particular interest is the scenario where customers do downsell. This is a more realistic scenario, and this provides revenue gains of 2.90 %. We therefore have found evidence that this framework is particularly beneficial in these cases.

6.6 DISCUSSION

The results show our ability to segment ODs in three distinct categories (price-, schedule- and comfort-sensitive). Currently, we use bookings from the entire period for which we have been given access to, as one big dataset. It could be that there is seasonality in what customers prefer. For example, in the summer peak, when schools have their holidays, people may be less price sensitive. After all, they are bound by these dates to travel, much more so than in an off-peak period, when it is easy to move your travels a day or two.

Another detail we would like to highlight is that the data shown here is from one specific OTA. This brings us to the following. Firstly, it could be that there are airlines that do not sell ticket through this OTA. For example, in the US, Southwest Airlines only sells tickets through their own channel, not through any OTA. Secondly, while we dealt with a large enough number of bookings to make our research statistically significant, one has to remember that the airline also sells through other channels, including other OTAs. For this reason, the dataset may not represent the entire population. However, we cannot think of reasons why customer behavior on one OTA will be drastically different than behavior on another channel.

When using ratings from airlines, we chose to calculate aggregates by airline, not by OD and airline. As discussed, we took this decision because the number of reviews would get extremely small and as a result aggregates are not reliable. Particularly for airlines that offer many different types of aircraft - some of them having certain features, such as WiFi or IFE, while other aircraft do not - this may not be fair to aggregate review scores like this.

While elements such as safety records may be a deciding factor intuitively, the number of accidents or incidents is at an historically low level. In effect, flying has become very safe, and therefore, there is little difference in measures between airlines. For this reason, we suspect the model is not using this feature for any OD.

Intra-day changes to fares can cause the fares from the booking dataset mismatch those from the competitive pricing dataset. While bookings depend on (real time) availability, the fares in the competitive pricing dataset are only scraped once a day (but always at the same time).

The implications for RM are as follows. We have shown that on some ODs, passengers only seem to look at fare in their decision-making process. As a result, the airline could considering lowering fares for these ODs. Alternatively, it could look into unbundling fares to become more competitive. On the other hand, the approach in this chapter extends to other departments of airlines: it shows that the investments some airlines make in their entertainment offering do drive extra bookings.

In traditional RM optimisation techniques, one optimizes network revenue using demand forecasts, estimated fares as well as capacity constraints. Some techniques assume booking class-independent demand forecasts, while others use customer choice probabilities to derive demand forecasts. In this chapter, we have presented a framework to estimate whether a given itinerary will get purchased or not. Naturally, rolling up these estimates from an itinerary to OD-level, we can derive a true, competitor-based demand forecast.

Another implication for RM is having an ability to estimate what your airline's fare premium should be, given your offering. It is often a discussion in airlines how much more expensive, or similarly, how much more cheaper you should be compared to your competitors given your product. For figure 6.13 we estimated what this airline's premium could be for this OD, given its very well-rated IFE offering.

Before discussing the results, we compared the logit to the XGB model. We found that the XGB model outperforms the logit model for true positives. However, we also found that the logit model is more biased toward predicting false negatives while the XGB model is more biased toward false positives. In practice, we would prefer a false positive to some extent. After all, predicting a false negative is arguably worse than a false positive: research typically shows that overforecasting results in better revenue performance than underforecasting. It may be interesting to study why these methods are biased toward different errors, but this is outside the scope of this chapter.

We argue that the most important result of our work is obtaining a better forecast. Weatherford et al [2], show that overforecasting by 25% accuracy lead to a revenue loss of 1.35%, while overforecasting by 12.5% lead to revenues 0.18% lower than an optimal forecast. The fares ranged from \$66 to \$275 in their work, with demand consistent across classes. Comparing our work to this work, we suspect that our results are higher as a result of a wider-spread fare ladder and a different optimization technique. Since our forecast considers competition, this could be a straightforward method to incorporate competitor-based information into the optimization process. One should remember that the results presented in this chapter, in particular in Table 6.17, are a lower bound to actual revenue performance. After all, only 42% of the flight's total demand was modelled using our new technique. The objective of this chapter was to introduce a new method to predict booking behavior. The results we showed have a demand factor of 0.98. We suggest further research into different scenarios - for example, extremely empty of extremely popular flights; different optimization techniques; different fare ladders to study the effects of our method on revenue performance.

Finally, the method used in this chapter is a black box method. As we discussed, while a decision tree is easy to understand, an ensemble of decision trees is not. Therefore, the analyst may have its doubts on how the model arrives at its prediction. The *xgboostExplainer* tool, which we used to produce Figures 6.11, 6.12 and 6.13 is a great tool to give insight in the model, which in turn will restore analyst confidence.

6.7 CONCLUSION

In summary, we have the following observations and conclusions:

1. Traditional forecasting methods tend to only use historical observations of demand, and do not consider the *why*.
2. To understand *why* people make a decision, data from competitor pricing, airline ratings (crew, food and beverage, in-flight entertainment, and more), social media sentiment, safety ratings and fleet information is combined. From these, features are engineered.
3. Next, we study how the host airline and competitor airlines fare's move over time. Features are engineered that compare the current fare against the fare at the time of booking.
4. Using extreme gradient boosting, we show that OD-pairs can be categorized into price-sensitive (driven by how fares have moved over time), schedule-sensitive (driven by how desirable a schedule is) and comfort-sensitive (driven by the quality of the in-flight entertainment).
5. Most interestingly, we found that the quality of in-flight entertainment was found very important, but only on long-distance trips.
6. Intriguingly, features such as social media, safety ratings and fleet information were not deemed important.
7. Through simulation, we show that this method of forecasting outperforms the most commonly-used forecasting method and show improvements between 0.7% and 2.9%.

6.8 APPENDIX: ENGINEERED FEATURES

Field Title	Field Description	Source	Calculation
airline.x	Airline	Competitor Pricing	
od	OD	Competitor Pricing	
airline_num_id	Obfuscated Airline ID	Schedule	
num	Obfuscated date	OTA	
t.x	Days before departure	OTA	
home_carrier	Is this airline a home carrier?	Schedule	Is OD's origin or destination the airline's hub?
sc1	Review score site 1	Review Website	
sc2	Review score site 2	Review Website	
price	Price of itinerary	Competitor Pricing	
rating_l_recommended	Airline recommended by Leisure Passengers	Review Website	Median rating of all review
rating_l_review	Review sentiment by Leisure Passengers	Review Website	Text mining based on AFINN dataset
rating_l_fb	F&B rating by Leisure Passengers	Review Website	Median rating of all review
rating_l_ground	Ground services rating by Leisure Passengers	Review Website	Median rating of all review
rating_l_ife	IFE rating by Leisure Passengers	Review Website	Median rating of all review
rating_l_crew	Crew rating by Leisure Passengers	Review Website	Median rating of all review
rating_l_seat	Seat rating by Leisure Passengers	Review Website	Median rating of all review
rating_l_value	Value for money rating by Leisure Passengers	Review Website	Median rating of all review
rating_l_wifi	WiFi rating by Leisure Passengers	Review Website	Median rating of all review
rating_l_obs	Number of review observations by Leisure Passengers	Review Website	Count of number of reviews
rating_b_recommended	Airline recommended by Business Passengers	Review Website	Median rating of all review
rating_b_review	Review sentiment by Business Passengers	Review Website	Text mining based on AFINN dataset
rating_b_fb	F&B rating by Business Passengers	Review Website	Median rating of all review
rating_b_ground	Ground services rating by Business Passengers	Review Website	Median rating of all review
rating_b_ife	IFE rating by Business Passengers	Review Website	Median rating of all review
rating_b_crew	Crew rating by Business Passengers	Review Website	Median rating of all review
rating_b_seat	Seat rating by Business Passengers	Review Website	Median rating of all review
rating_b_value	Value for money rating by Business Passengers	Review Website	Median rating of all review
rating_b_wifi	WiFi rating by Business Passengers	Review Website	Median rating of all review
rating_b_obs	Number of review observations by Business Passengers	Review Website	Count of number of reviews
sent_mean	Mean sentiment score	Review Website	Median rating of all review
sent_sd	Standard deviation sentiment score	Review Website	Text mining based on AFINN dataset
sent_mean_rel_diff	Difference to mean sentiment score	Review Website	Median rating of all review
sent_mean_rel_perc	Percentage difference to mean sentiment score	Review Website	Median rating of all review
sent_sd_rel_diff	Difference to sd sentiment score	Review Website	Median rating of all review
sent_sd_rel_perc	Percentage difference to sd sentiment score	Review Website	Median rating of all review
direct_flight	Is this a direct flight yes/no?	Review Website	Median rating of all review
has_night_flight	Does this airline offer a night flight?	Review Website	Median rating of all review
has_day_flight	Does this airline offer a day flight?	Review Website	Median rating of all review
first_flight_dep	Airline's time of departure of first flight of the day	Review Website	Count of number of reviews
first_flight_arr	Airline's time of arrival of first flight of the day	Review Website	Median rating of all review
last_flight_dep	Airline's time of departure of last flight of the day	Review Website	Text mining based on AFINN dataset
last_flight_arr	Airline's time of arrival of lasst flight of the day	Review Website	Median rating of all review
min_flying_time	Airline's minimum flying time	Review Website	Median rating of all review
min_conn_time	Airine's minimum connection time	Review Website	Median rating of all review
min_travel_time	Airline's minimum travel time	Review Website	Median rating of all review
has_night_departure	Is this a night departure?	Review Website	Median rating of all review
has_morning_arrival	Is this a morning arrival?	Review Website	Median rating of all review
num_frequencies	Number of frequencies offered by airline	Review Website	Median rating of all review
aircraft_type	Aircraft type	Review Website	Count of number of reviews
airline_fleet_size	Airline fleet size	Kaggle	Sum of airframes
airline_fleet_cost	Airline fleet cost (estimated)	Kaggle	Sum of airframe cost
airline_fleet_age	Airline fleet age	Kaggle	Average age of airframes
bucket_t	Bucketed time (time before departure grouped in multiples of 10)	OTA	floor(Time before departure/10)*10

Field Title	Field Description	Source	Calculation
mean3d_al	Mean last 3 day price moment of own airline	Competitor Pricing	Mean of fare at t.x - fare at t.y, rolling 3 days, airline vs itself
min3d_al	Min of last 3 day price moment of own airline	Competitor Pricing	Min of fare at t.x - fare at t.y, rolling 3 days, airline vs itself
max3d_al	Max of last 3 day price moment of own airline	Competitor Pricing	Max of fare at t.x - fare at t.y, rolling 3 days, airline vs itself
sd3d_al	SD of last 3 day price moment of own airline	Competitor Pricing	SD of fare at t.x - fare at t.y, rolling 3 days, airline vs itself
mean7d_al	SD of last 7 day price moment of own airline	Competitor Pricing	Mean of fare at t.x - fare at t.y, rolling 7 days, airline vs itself
min7d_al	SD of last 7 day price moment of own airline	Competitor Pricing	Min of fare at t.x - fare at t.y, rolling 7 days, airline vs itself
max7d_al	SD of last 7 day price moment of own airline	Competitor Pricing	Max of fare at t.x - fare at t.y, rolling 7 days, airline vs itself
sd7d_al	SD of last 7 day price moment of own airline	Competitor Pricing	SD of fare at t.x - fare at t.y, rolling 7 days, airline vs itself
mean14d_al	SD of last 14 day price moment of own airline	Competitor Pricing	Mean of fare at t.x - fare at t.y, rolling 14 days, airline vs itself
min14d_al	SD of last 14 day price moment of own airline	Competitor Pricing	Min of fare at t.x - fare at t.y, rolling 14 days, airline vs itself
max14d_al	SD of last 14 day price moment of own airline	Competitor Pricing	Max of fare at t.x - fare at t.y, rolling 14 days, airline vs itself
sd14d_al	SD of last 14 day price moment of own airline	Competitor Pricing	SD of fare at t.x - fare at t.y, rolling 14 days, airline vs itself
mean28d_al	SD of last 28 day price moment of own airline	Competitor Pricing	Mean of fare at t.x - fare at t.y, rolling 28 days, airline vs itself
min28d_al	SD of last 28 day price moment of own airline	Competitor Pricing	Min of fare at t.x - fare at t.y, rolling 28 days, airline vs itself
max28d_al	SD of last 28 day price moment of own airline	Competitor Pricing	Max of fare at t.x - fare at t.y, rolling 28 days, airline vs itself
sd28d_al	SD of last 28 day price moment of own airline	Competitor Pricing	SD of fare at t.x - fare at t.y, rolling 28 days, airline vs itself
mean3d_yy	Mean last 3 day price moment of cheapest airline	Competitor Pricing	Mean of fare at t.x - fare at t.y, rolling 3 days, airline vs cheapest airline
min3d_yy	Min of last 3 day price moment of cheapest airline	Competitor Pricing	Min of fare at t.x - fare at t.y, rolling 3 days, airline vs cheapest airline
max3d_yy	Max of last 3 day price moment of cheapest airline	Competitor Pricing	Max of fare at t.x - fare at t.y, rolling 3 days, airline vs cheapest airline
sd3d_yy	SD of last 3 day price moment of cheapest airline	Competitor Pricing	SD of fare at t.x - fare at t.y, rolling 3 days, airline vs cheapest airline
mean7d_yy	SD of last 7 day price moment of cheapest airline	Competitor Pricing	Mean of fare at t.x - fare at t.y, rolling 7 days, airline vs cheapest airline
min7d_yy	SD of last 7 day price moment of cheapest airline	Competitor Pricing	Min of fare at t.x - fare at t.y, rolling 7 days, airline vs cheapest airline
max7d_yy	SD of last 7 day price moment of cheapest airline	Competitor Pricing	Max of fare at t.x - fare at t.y, rolling 7 days, airline vs cheapest airline
sd7d_yy	SD of last 7 day price moment of cheapest airline	Competitor Pricing	SD of fare at t.x - fare at t.y, rolling 7 days, airline vs cheapest airline
mean14d_yy	SD of last 14 day price moment of cheapest airline	Competitor Pricing	Mean of fare at t.x - fare at t.y, rolling 14 days, airline vs cheapest airline
min14d_yy	SD of last 14 day price moment of cheapest airline	Competitor Pricing	Min of fare at t.x - fare at t.y, rolling 14 days, airline vs cheapest airline
max14d_yy	SD of last 14 day price moment of cheapest airline	Competitor Pricing	Max of fare at t.x - fare at t.y, rolling 14 days, airline vs cheapest airline
sd14d_yy	SD of last 14 day price moment of cheapest airline	Competitor Pricing	SD of fare at t.x - fare at t.y, rolling 14 days, airline vs cheapest airline
mean28d_yy	SD of last 28 day price moment of cheapest airline	Competitor Pricing	Mean of fare at t.x - fare at t.y, rolling 28 days, airline vs cheapest airline
min28d_yy	SD of last 28 day price moment of cheapest airline	Competitor Pricing	Min of fare at t.x - fare at t.y, rolling 28 days, airline vs cheapest airline
max28d_yy	SD of last 28 day price moment of cheapest airline	Competitor Pricing	Max of fare at t.x - fare at t.y, rolling 28 days, airline vs cheapest airline
sd28d_yy	SD of last 28 day price moment of cheapest airline	Competitor Pricing	SD of fare at t.x - fare at t.y, rolling 28 days, airline vs cheapest airline
mean3d_zz	Mean last 3 day price moment of second cheapest airline	Competitor Pricing	Mean of fare at t.x - fare at t.y, rolling 3 days, airline vs second cheapest airline
min3d_zz	Min of last 3 day price moment of second cheapest airline	Competitor Pricing	Min of fare at t.x - fare at t.y, rolling 3 days, airline vs second cheapest airline
max3d_zz	Max of last 3 day price moment of second cheapest airline	Competitor Pricing	Max of fare at t.x - fare at t.y, rolling 3 days, airline vs second cheapest airline
sd3d_zz	SD of last 3 day price moment of second cheapest airline	Competitor Pricing	SD of fare at t.x - fare at t.y, rolling 3 days, airline vs second cheapest airline
mean7d_zz	SD of last 7 day price moment of second cheapest airline	Competitor Pricing	Mean of fare at t.x - fare at t.y, rolling 7 days, airline vs second cheapest airline
min7d_zz	SD of last 7 day price moment of second cheapest airline	Competitor Pricing	Min of fare at t.x - fare at t.y, rolling 7 days, airline vs second cheapest airline
max7d_zz	SD of last 7 day price moment of second cheapest airline	Competitor Pricing	Max of fare at t.x - fare at t.y, rolling 7 days, airline vs second cheapest airline
sd7d_zz	SD of last 7 day price moment of second cheapest airline	Competitor Pricing	SD of fare at t.x - fare at t.y, rolling 7 days, airline vs second cheapest airline
mean14d_zz	SD of last 14 day price moment of second cheapest airline	Competitor Pricing	Mean of fare at t.x - fare at t.y, rolling 14 days, airline vs second cheapest airline
min14d_zz	SD of last 14 day price moment of second cheapest airline	Competitor Pricing	Min of fare at t.x - fare at t.y, rolling 14 days, airline vs second cheapest airline
max14d_zz	SD of last 14 day price moment of second cheapest airline	Competitor Pricing	Max of fare at t.x - fare at t.y, rolling 14 days, airline vs second cheapest airline
sd14d_zz	SD of last 14 day price moment of second cheapest airline	Competitor Pricing	SD of fare at t.x - fare at t.y, rolling 14 days, airline vs second cheapest airline
mean28d_zz	SD of last 28 day price moment of second cheapest airline	Competitor Pricing	Mean of fare at t.x - fare at t.y, rolling 28 days, airline vs second cheapest airline
min28d_zz	SD of last 28 day price moment of second cheapest airline	Competitor Pricing	Min of fare at t.x - fare at t.y, rolling 28 days, airline vs second cheapest airline
max28d_zz	SD of last 28 day price moment of second cheapest airline	Competitor Pricing	Max of fare at t.x - fare at t.y, rolling 28 days, airline vs second cheapest airline
sd28d_zz	SD of last 28 day price moment of second cheapest airline	Competitor Pricing	SD of fare at t.x - fare at t.y, rolling 28 days, airline vs second cheapest airline
dep_time_mam	Departure time in minutes after midnight	Schedule	Hours and minutes converted into minutes
connecting_time	Connecting time	Schedule	Dep time next flight - Arr time previous flight
travel_time	Travel time	Schedule	Flying time + connecting time
mintt	Minimum connecting time offered by airline	Schedule	Min(travel_time) by OD by days before departure
tt_delta	Difference between airline travel time and min travel time for this OD	Schedule	Airline's travel time - mintt
dept_delta	Difference between itinerary's departure time and "ideal" deparure time	Schedule	Departure time - 7AM departure
mkt_fare	Lowest fare in the market	Competitor Pricing	Min fare by OD by time before departure
mkt_fare_diff	Difference to lowest fare in the market	Competitor Pricing	Airline's fare - mkt_fare
mkt_fare_diff_perc	Percentage difference to lowest fare in the market	Competitor Pricing	(Airline's fare/mkt_fare)-1
is_cheapest	Is this airline cheapest?	Competitor Pricing	If airline's fare = mkt_fare
is_bought	Itineray purchased? Label	OTA	Is this itinerary bought?
airline_id	Obfuscated airline ID	OTA	Randon number

Table 6.18: Overview of all engineered features

7

MODELLING CANCELLATIONS FOR AIRLINE BOOKINGS

Abstract

Most of the research into forecasting in the field of RM concerns forecasting demand. Cancellations have not seen the exposure that they deserve. After all, cancellations pose two risks to the airline. First, customers cancelling shortly before flight departure do not allow the airline to resell these seats, as there is insufficient time. For this reason, it is important to estimate the time of cancellation. Second, a reservation takes up a unit of capacity, which typically increases price and therefore may turn other potential customers away. The research in the field of cancellations focuses on whether a reservation is cancelled, not at what time. In this chapter, we study five different processes, which we identified in Chapter 6. Next, we introduce a three-step framework: first, it is identified whether a reservation is likely to cancel. Second, the type of cancellation is identified. Third, the time of cancellation is estimated through a modified Geometric distribution with probabilities found using Bayesian inference.

This chapter is based on [21].

7.1 INTRODUCTION

Pricing is the science of segmenting the market and setting appropriate price levels. Determining how many products to sell of each of these price levels, through inventory, is the objective of RM. Forecasting demand is crucial for developing good policies. Given a segmented market, associated demand forecasts and price levels, these are then used as inputs to extract maximum revenues. In practice, forecasting demand may not be as straightforward as it seems. After all, *realized demand is the result of bookings that have not cancelled*. This calls for not only an accurate demand forecast, but an accurate cancellation forecast as well. However, just like people book for different reasons - some may prioritize price over schedule, while others may prioritize luggage allowance over travel time - people cancel for different reasons.

It is important to stress the difference between a cancellation and a no-show. We follow Van Ryzin's definitions [65], in that a cancellation "is a reservation that is withdrawn by a customer strictly prior to the time of service". On the other hand, a no-show "(is a reservation that) is not cancelled and does not show up in time". Note that for cancellations, there is time to resell a unit of capacity, while there is no such time for a no-show. This illustrates the importance of overbooking: selling more seats than physical capacity. In our work, we only focus on cancellations.

Understanding at what time before departure people cancel is important, since this allows the airline to make better decisions which passenger to accept on one hand, and to establish better overbooking levels on the other hand. To illustrate the dynamics of cancellations, suppose an airline is faced with the following problem: three days before departure, two customers arrive and it only has one seat left for sale. One customer intends to pay only $500 and does not have the ability to cancel. Another customer intends to pay $800, but has the ability to cancel. Traditionally, cancellation rates are estimated without the element of time: a booking is cancelled yes or no. Assume that this second customer has a probability of cancellation of 0.5. Clearly, the expected value of this customer is $400, less than that of the first customer. However, what if we expect this customer to cancel within the same day of booking? This will give the airline the opportunity to sell that unit of capacity again, perhaps attracting a customer who is willing to pay $800, but does not intend to cancel. Of course, if we expect the customer to cancel right before departure, there is no sufficient time left to sell, the empty seat is worthless, and we should have accepted the first customer.

While estimating the time of cancellation is important and can give insight in customer behavior, this should not be the objective itself. The true importance of this cancellation forecast can be studied by using this model in an optimization problem, that is aimed at developing policies that maximize revenue, given these cancellation forecasts.

This chapter is organized as follows. A discussion of the topic of cancellations in given in Section 7.2. Section 7.3 provides a data analysis of the data that was supplied to us and is followed by our framework and its modelling. We discuss the performance of our model in Section 7.4. In Section 7.5 we discuss our work and provide suggestions for future research. We put forth conclusions in Section 7.6.

7.2 REASONS FOR CANCELLATIONS

Cancellations happen for different reasons. After discussing with subject-matter experts, we have identified five possible reasons for cancellations, along with another category which covers all remaining reasons. This results into six different groups:

1. **F**are of booked **a**ir**l**ine ($Fal(t)$): fares constantly move up and down for your own airline. "Own airline" is defined as the "host" airline: this research is written from the perspective of this airline. These may influence a customer's cancellation behavior. This value captures the fare of the airline at time t.

2. **Av**ai**l**ability ($Avl(t)$): similar to how fares move up and down, availability may change over time. We have covered the danger of this in Chapter 6. This could result in a customer cancelling and rebooking. $Avl(t)$ denotes the current availability at time t.

3. **F**are of a competitor, in the **m**ar**k**e**t** ($Fmkt(t)$): similar to fares of the airline booked, fares of competitors (using the terminology introduced above, the host airline is the "own" airline, while its competitor airlines in the market may also be called "target" airlines) change. This could result in a customer changing their mind and cancelling the itinerary booked. This variable captures the fare of the cheapest competitor at time t.

4. **Sch**edule ($Sch(t)$): airlines change schedules from time to time. A change in schedule could result in a product no longer be acceptable to a customer. This feature captures the airline's current schedule in minutes after midnight.

5. **Alt**ernative offering ($Alt(t)$): In Chapter 6, we model why customers book a given itinerary. A better alternative may become available, which may result in cancellation of a currently booked itinerary. This metric captures the likelihood of purchase for alternative offers at time t.

6. Noise ($Noise(t)$): this category covers all other reasons for cancellations. We introduce this category so we can classify all cancellations. This variable measures the noise at time t.

The objective is, given these parameters, to estimate the probability that a booking is cancelled at some time τ. However, having these parameters available at the current point of time is not enough. Clearly, this should be in relation to what these parameters were *when the itinerary was originally booked*. For example, consider an itinerary that is currently offered at $200. Now consider two customers: one that has purchased this itinerary at $1000, and another customer who paid $250. Arguably, the probability that the former customer cancels is higher than the latter customer, since the fare difference between the current and originally purchased fare is much greater. Similarly, other features should be considered in relation to those at time of booking. Note that $Fal(t)$, $Avl(t)$, $Fmkt(t)$, $Sch(t)$ and $Alt(t)$ all denote events. These are random variables, depending on time t: after all, it is not a-priori known whether, for example, a better schedule will become available at some t.

Let t_b be the time of booking in terms of day before departure and define t_c, the current time in terms of days before departure. We then introduce:

$$\Delta(X) = X(t_c) - X(t_b); \text{for } X \in \{\text{ Fal, Avl, Fmkt, Sch, Alt }\} \quad (7.1)$$

While $Fal(t)$, $Avl(t)$, $Fmkt(t)$, $Sch(t)$ and $Alt(t)$ are events, the function in Equation 7.1 measure *what type* of event happened. These five functions measure how each metric has moved, since time of booking up to the current time t_c. $\Delta(Fal)$ and $\Delta(Fmkt)$ measure how fares have moved. $\Delta(Avl)$ represents the number of classes. $\Delta(Alt)$ measures how probabilities have changed for competing itineraries. Finally, $\Delta(Sch)$ measures how the schedule has changed in terms of minutes between t_c and t_b.

In practice, airlines themselves may not have access to all data sources. For example, airlines have historical fares stored as this is required for pricing engines. However, it may not have access to historical fares of competitors. Historical schedules are often present; these are, by definition, leg-level measures and as such do not require large storage space. On the other hand, historical availability is often not stored as these are POS/OD-specific, and therefore, require large amounts of space.

It should be noted that these categories are not always mutually exclusive: for example, in Chapter 6, we show that price can have great influence on how popular an itinerary is, particularly on short-haul itineraries.

7.3 MODELLING

In Chapter 6, we studied the five different processes introduced in the section above and showed how these affect the likelihood of an itinerary being purchased. In particular, we see that changes over time impact a customer's decision. For example, if an airline was not the cheapest in the market, but then lowers its fare and becomes the cheapest, we show the likelihood of purchase increases dramatically.

Similarly, we hypothesize that these processes also affect whether or not a booking is cancelled. For cancellations, we introduce two other processes over time: a process that describes *ticketing*, and one that describes the *time of departure.*

Ticketing

A booking is created when a customer *intends* to travel. Once a customer has paid, a booking is said to be *ticketed*, and a customer is cleared to fly. The time between booking and ticketing is called the *ticketing time*. This time exposes the airline to risk, as a customer is taking up a unit of capacity but has no obligation to pay until the ticketing time limit. These ticketing time limits often depend on the booking class, on the time before departure and booking channel. An example of these are given in Tables 7.1 and 7.2 below. For an itinerary with a departure date one week from now, payment needs to be completed within three days from the moment of booking. On the other hand, for a booking departing six months from now, ticketing must be completed two months before departure.

ADVANCE RESERVATIONS	TICKETING MUST BE COMPLETED WITHIN 72 HOURS AFTER RESERVATIONS ARE MADE

Table 7.1: Rules for a ticket departing one week away

ADVANCE RESERVATIONS	TICKETING MUST BE COMPLETED 60 DAYS BEFORE DEPARTURE OF THE FIRST INTERNATIONAL SECTOR

Table 7.2: Rules for a ticket departing six months away

These restrictions may also differ from point-of-sale to point-of-sale, and may be different across fare classes. However, all of these are known and can be found in the fare rules.

Departure time

The process up to the departure time describes the behavior of cancellation in the final seven days, $t = T-6, T-5, ..., T$. This is a number we arbitrarily chose: cancellations that happen outside this window (that is, seven days before departure or earlier) are assumed not to have a relation to cancellation behavior.

Estimating *Fal*

Fal follows the process of how the airline's fare moves over time since the time of booking. The process does not use the actual fare, but rather whether the fare difference exceeds some $\epsilon > 0$ since time of booking. Let f_b be the fare paid by a customer at the time of booking t_b, and let f_c be the current fare for this product at the current time t_c. Therefore, the variable *Fal* may take on a value of 1 if $f_c > f_b$, 0 if $abs(f_c - f_b) < \epsilon$ or equal to -1 if $f_c - f_b < -\epsilon$. Since the value of f_b is given, the only unknown is f_t for future times t. In the next section, we will discuss how to reliably estimate *fal*. We also found that other events (namely, *fmkt, sch, avl and alt*) do not have a significant impact on cancellations. For this reason, we only discuss *fal*.

Estimating event times

The objective of modelling these departure times is to use this information to predict when a cancellation may occur. However, a-priori, we do not know when events presented in Equation (7.1) will happen. To use these processes to predict when a product is cancelled, we would first need to predict when an event happens. This may be challenging. However, for the fal, tkt and dep events, we are able to provide estimates: the fare rules specify when the tkt event will happen. And, by definition, dep always occurs at $t = 0$. Typically, fares are reviewed periodically, every n weeks, so a reliable estimate for the next fal event may be obtained. Having obtained estimates of what values of t these events happen, we study the days after the fal event and the days leading up to tkt and dep events.

7.3.1 *Framework*

In this section, we present our framework to estimate cancellation times is introduced. Traditionally, cancellations are forecasted by product (typically OD-pair, POS, class) and data collection point (DCP, a range of time in the booking curve, to lower the variability; rather than recording the number of observations recorded in a single day). To accomplish this, the number of cancellations is counted in each time interval before the number of cancellations is predicted by any statistical method. The rate of cancellation is then averaged within the time interval. For example, if a given interval is of length 14 days, and the cancellation forecast is 7, a cancellation rate of 1/2 per day is expected.

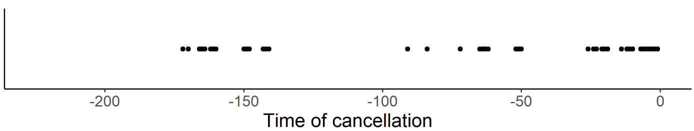


Figure 7.1: Example of cancellation times

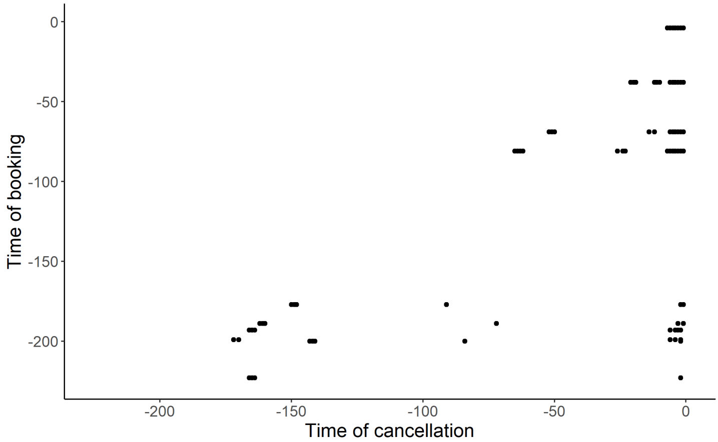


Figure 7.2: Example of cancellation times, by time of booking

Consider Figure 7.1. Every point represents a cancellation recorded at some time t_c before departure. Traditionally, this is the data that is used to predict. There is an inherent danger with this approach: a cancellation can not occur without a booking. Refer to Figure 7.2. There, the same data is used, but we now show the time of booking t_b on the vertical axis. Clearly, fixing the time of booking, we identify two types of cancellations: those made shortly *after booking* (which appear just to the right of the diagonal) and those made shortly *before departure* (which appear to the far right). We will use this insight later. To be clear: taking the sum over all t_b will generate Figure 7.1. Consider the two cancellations shown at $t_c = -27$ in Figure 7.1. These cancellations are a result of a single time of booking, t_b.

Figure 7.3 shows the distribution of time of cancellation t_c for ten different booking dates t_b. We note two things. First, we observe an increasing probability of cancellation at some time after booking. What follows, is a period of time with low probability mass, before this probability increases again before the time of departure $t = 0$. Note that this pattern seems to be consistent across different values of t_b, but the mass is moving from shortly after booking to shortly before departure. For example, compare $t_b = -90$ with $t_b = -23$. Both exhibit the aforementioned increasing probability shortly after booking, as well as before time of departure $t = 0$, but $t_b = -90$ has more mass closer after booking while $t_b = -23$ has more mass close to departure. Second, observe that the time between t_b and $t = 0$ exhibits low probabilities of cancellation across values of t_b.

While there is mass for t_c for all values of t_b, it is evident that the distribution of t_c depends on t_b. However, from this figure it seems there is no linear relationship: the values of t_c should lie close to t_b, close to $t = 0$, or between these values with a small probability.

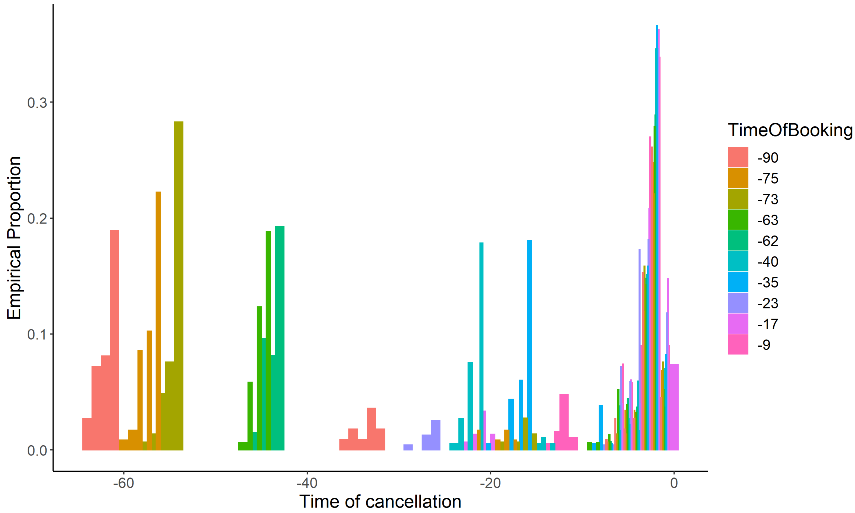


Figure 7.3: Cancellation times

Intuitively, this makes sense: a customer may realize a mistake in travel dates once their ticket is received. On the other hand, cancelling right before departure may indicate a customer's plans changed last minute, or has simply forgotten about their flight. As we discussed in [17], these can be considered "smart customers", as there is no financial incentive to cancel early.

Therefore, we first estimate whether a booking has potential to be cancelled early or not. Note that at this stage we do not estimate whether a booking actually cancels, we only classify a booking as a *potential cancellation.*

From historical data, we can see that there is a time lag between the airline (or market) dropping their fares and the cancellation. However, there is a clear increase in cancellation probability. Suppose a fare change will happen at time t_c. Then the final probability we estimate will have an informative prior to reflect this. Predicting t_c at first hand may seem complicated, but recent work by Fiig et al. [182] has shown that predicting pricing can be attained within 10% accuracy. Moreover, keep in mind that this variable we use, *fal*, is one's own (host) airline fare behavior, and therefore, internal information can be exploited. Typically, pricing teams do fare reviews at fixed intervals (for example, once every week for highly competitive ODs, once every quarter for ODs with low demand) and change fares if required. This is typically done by benchmarking against competitors in the market. This also reaffirms our statement of the variable *Fal* being a proxy for *Fmkt*.

7.3.2 *Framework*

In this section, we will introduce our framework to predict the time of cancellation. To this end, define the following notation:

- t_b is the time of booking (given);
- C_j is a (binary) random variable that represents whether product j will cancel (independent of time);
- $T_j(t)$ is a random variable that defines the time product j will cancel (to be estimated);
- $\Theta_j(t_b)$ is random variable that specifies the type of cancellation, driven by ticketing, fare movement or departure time.

Therefore, we suggest the following stepwise approach:

Step 1: Calculate the probability dat a product j will cancel, $P(C_j = 1)$. This is discussed in Section 7.3.3.

Step 2: Given that a booking will cancel, estimate what type of cancellation we expect through the distribution of $\Theta_j(t_b)$. This is covered in Section 7.3.4.

Step 3: Calculate the time of cancellation $P\big(T_j(t) = t_b\big)$ for values $t_b, t_{b+1}, \ldots$. This is illustrated in Section 7.3.6.

7.3.3 *Step 1: determining cancellation*

In this section, we will estimate $P(C_j = 1)$: the probability that an itinerary is cancelled (irrespective of time of cancellation). This is the most common part of research. One can estimate C_j in two different ways. It can be done by means of an aggregated way, by using counts of each product j, before making predictions and forecasts based on these (aggregated) data points. Alternatively, each record can be used as an individual data point, predicted whether cancels, and rolled up to an aggregate. This latter approach is usually referred to as a PNR (*passenger name record*, or *booking*)-model. I have discussed these differences in Chapter 2. In this section, we use a PNR-approach to predict the cancellation probability.

At the time of booking, in absence of any frequent flyer data, we have limited data about a customer. Even if a customer has a frequent flyer account and is logged in, most passengers make a journey at most once a year, so there will be very limited historical data to predict future behavior. Introduce the following variables:

1. r_i is the booking class purchased. In our dataset, we have $i = 1, .., 12$. As per convention, r_1 is the most expensive class, while r_{12} the cheapest;
2. a_i is the availability (the lowest available class for which at least one seat is available) at the time of booking;
3. l is an indicator, which is equal to 1 if a cheaper class was available for sale at the time of purchase:
$$l = \begin{cases} 1 & \text{if } r_i > a_i \\ 0 & \text{otherwise;} \end{cases}$$
4. s is the number of searches for this specific itinerary;
5. ρ an indicator, equal to 1 if the airline was cheapest in the market at time of purchase (fal).

Note that r_i can be directly observed from reservation data. The variable s was obtained from website analytics. Parameters ρ and ϕ can be obtained from a price benchmarking dataset, which captures daily airfare and schedule. a_i and I can be obtained by using availability data. All these variables, except s, can be obtained from the public domain, but many airlines use commercial solutions.

	Dependent variable:	
	Cancellation	
	Logit	*Xgboost*
	(1)	(2)
Days Prior, t_b	-0.001^{***}	1
	(0.00005)	
Booking Class, r_i	-0.039^{***}	3
	(0.001)	
# Searches, s	-0.001	2
	(0.003)	
Lower class available, $I_{r_i > a_i}$	0.503^{***}	5
	(0.007)	
Cheaper airline available, Fal	0.482^{***}	4
	(0.006)	
Constant	0.125^{***}	–
	(0.012)	
Observations	313	313
R^2	0.553	
Adjusted R^2	0.552	
# Iterations		500
Training Error Improvement		1.5%
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 7.3: Model results - logistic regression and XGB. The value given is the coefficient, the value within parentheses the standard error. For XGB, we note the feature importance based on information gain.

Table 7.3 shows the model parameters and statistical significance of logistic regression and gradient boosting.

7.3.4 *Step 2: determining cancellation type*

In this section, we propose a heuristic to determine what type of cancellation we expect. These types follow from the processes introduced in Section 7.3. To accomplish this, we build three different, independent, models that predict whether a given product is likely to be affected by this process. The output of these models is the probability that a product j booked at time t_b is of type **T**, **D** or **F**. These probabilities follow from a $Bernoulli(\theta_j^T(t_b))$, $Bernoulli(\theta_j^D(t_b))$ and $Bernoulli(\theta_j^F(t_b))$ distribution. Three different logistic regression models were built to obtain the values for θ. Details of these logistic regression models are given in Table 7.4.

Let $p_j^T(t_b)$, $p_j^D(t_b)$ and $p_j^F(t_b)$ be the probabilities that product j are affected by the **T**, **D** and **F** process, respectively. It is then determined that product j are of the type that produces the largest probability. Define this type as $\Theta(j, t_b)$. Specifically,

$$\text{Type of cancellation for product } j \text{ booked at time } t_b = \Theta(j, t_b)$$
$$= \underset{T,D,F}{\arg\max} \left(p_j^T(t_b), p_j^D(t_b), p_j^F(t_b) \right)$$
$$\Theta(j, t_b) \in T, D, F \tag{7.2}$$

Let $\phi(T)$ be the count of products j assigned to type T, $\phi(D)$ be the count of products assigned to type D and $\phi(F)$ be the count of products assigned to type F. Therefore, we have:

$$\sum_{j=1}^{J} C_j = \phi(T) + \phi(D) + \phi(F) \tag{7.3}$$

Equation (7.2) ensures that every product j is assigned one type of cancellation, regardless of how low the probabilities p_j are. This ensures that Equation (7.3) is true: the number of cancellations, defined in Step One in Section 7.3.3, is equal to the sum over the number of type of cancellations.

	Dependent variable:		
	Ticketing	Departure Date	Fare Movement
	(1)	(2)	(3)
Days Prior, t_b	0.0003***	0.001***	−0.001***
	(0.0001)	(0.0001)	(0.0001)
Booking Class, r_i	0.090***	−0.084***	−0.007**
	(0.002)	(0.002)	(0.003)
# Searches, s	−0.107***	0.118***	−0.011
	(0.005)	(0.007)	(0.009)
Lower class available, $I_{r_i > a_i}$	0.0004	−0.001	0.001
	(0.009)	(0.010)	(0.014)
Cheaper airline available, Fal	−0.003	0.006	−0.003
	(0.008)	(0.009)	(0.013)
Constant	0.185***	0.720***	0.095**
	(0.025)	(0.031)	(0.042)
Observations	2,068	2,068	2,068
R^2	0.890	0.827	0.042
Adjusted R^2	0.890	0.826	0.039
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 7.4: Model results - ticketing, fare movement and departure date. The value given is the coefficient, the value within parentheses the standard error.

7.3.5 *Modelling effect of time*

In this section, we investigate how to calculate the probability of cancellation at time i: $P(X_i = 1 | X_j, \ldots, X_k = 0) \sim Bernoulli(\zeta)$ with $\forall j, k < i$. Naturally, that means it has not been cancelled yet, but we explicitly include this. We will use Bayesian Inference to model and update our belief of the unknown parameter ζ. We obtain these distributions for days around events. For example, suppose an event happens at time ϵ. Then we obtain ζ's for the range of ϵ, $\epsilon + 1$, $\epsilon + 2$; ζ_ϵ, $\zeta_{\epsilon+1}$, $\zeta_{\epsilon+2}$ ... The number of ζ's will be depend on the different processes that we model. We assume a hard cut off that depends on the process, after which we assume this process no longer has an affect on cancellation behavior.

In this section, we use a $Beta(\alpha = 1, \beta = 1)$ distribution for our prior distribution (the reason for our choice of parameters $\alpha = \beta = 1$ will become evident later in this section). The Beta distribution is defined on [0,1] and its density is given by:

$$P(\zeta) = \frac{\zeta^{\alpha-1}(1-\zeta)^{\beta-1}}{B(a,b)} \tag{7.4}$$

where $B(a,b)$ is a normalizing constant such that the above is a proper probability distribution function (that is, it integrates to one):

$$\int_0^1 \zeta^{\alpha-1}(1-\zeta)^{\beta-1} d\zeta \tag{7.5}$$

This is also known as the Beta function. It is known that the Beta distribution is a conjugate prior, so we have:

$$\begin{aligned} p(\zeta|y) &\propto \zeta^r(1-\zeta)^{n-r}\zeta^{\alpha-1}(1-\zeta)^{\beta-1} \\ &= \zeta^{r+\alpha-1}(1-\zeta)^{n-1+\beta-1} \\ &= Beta(\alpha+r, \beta+n-r), \end{aligned}$$

where $\propto$ represents "proportional to".

Therefore, the posterior distribution becomes:

$$p(\zeta|y) \sim Beta(y+1, n-y+1). \tag{7.6}$$

This classic result makes the process of Bayesian inference easier: rather than explicitly calculating the posterior distribution (7.4) after new data is obtained, we may simply update the parameters of the Beta distribution.

7.3.6 *Step 3: determining time of cancellation*

Consider the geometric distribution. This distribution models the process of having a number of failures before obtaining a success. The probability density function of being successful at the kth attempt is given by:

$$P(X = k) = (1-p)^{k-1}p, k = 1, 2, ..., \tag{7.7}$$

where X is the number of failures between successes.

To be consistent with the definition in Equation (7.7) above, we consider a non-cancellation a failure and a cancellation a success. Once a booking is made, we assume this booking follows a certain process. In this process, we discretize time to day-level: every day, we observe whether a booking is cancelled during this day, or whether it stays alive. If a booking is not cancelled, we record this as a failure, if a booking is cancelled, this is considered a success. In Section 7.3.1, we have seen that the probability of cancelling changes over time: the probability of cancellation between day 29 and 30 is different than the probability between day 30 and 31. This probability p depends on t. In short, Equation (7.7) cannot be modelled with a fixed value of p.

Introduce the probability $p_{t,j}(\tau)$, which is the probability that product j booked at time t is cancelled at time τ. Next, define $c_j(t,\tau)$ as the probability that product j booked at time t cancels at time τ. Then:

$$c_j(t,\tau) = p_{t,j}(\tau) \prod_{i=t}^{\tau-1} (1 - p_{t,j}(t_i)), \quad \tau > t. \tag{7.8}$$

In this formulation, a modified version of the geometric distribution of 7.7, the probability $p(t_i)$ now depends on time t. These values of $p(t_i)$ depend on the process identified, determined by Θ, and time t_i, as put forth in the framework in the previous sections.

7.4 RESULTS

In Section 7.4 we discuss our results. We first look at estimating whether a booking will cancel, estimating C_j, in Section 7.4.1. This is followed by Section 7.4.2, which evaluates the model's performance. Finally, Section 7.4.3 brings everything together and looks into our estimates for cancellation times.

7.4.1 *Estimating probability of cancellation*

Figure 7.4 shows the confusion matrices for the logistic regression (left) and gradient boosting method (right). We show an example of one particular product j. This dataset contained a total of $n = 10313$ bookings, of which $n_c = 3722$ (36.1%) were cancelled.

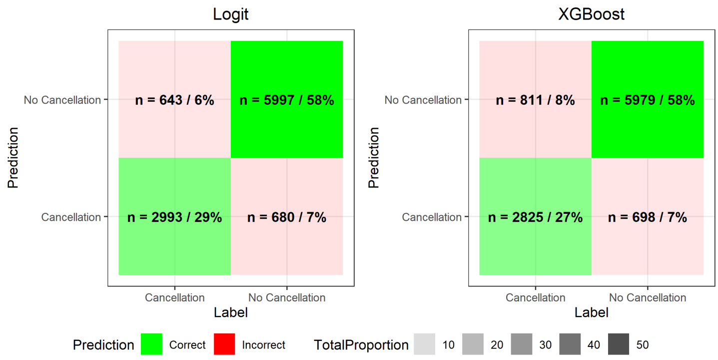


Figure 7.4: Cancellation performance by model, $n = 10313$

Consider the confusion matrix of the logistic regression model. Our model, with significant values shown in Table 7.3, achieves an accuracy of 87.2%. We also note that this regression is slightly biased toward false positives (7%), while false positives are predicted in 6% of all cases.

Model	Precision	Recall
Logistic Regression	0.81	0.82
Extreme Gradient Boosting	0.80	0.78

Table 7.5: Precision and recall of the logistic regression and extreme gradient boosting model

The precision and recall are given in Table 7.5. The precision, the ratio of true positives and true positives and false negatives, show that the logistic regression is 81% accurate in finding cancellations. The recall, the ratio of true positives and true positives and false positives, shows that 82% of those labelled cancelled, were actually cancelled.

Comparing the logistic regression with the XGB model, we report similar results, but particularly report more false positives. This results in an accuracy of 85.4%. Comparing recall with precision for the XGB model, it seems that the model is better at accurately predicting cancellations than finding the number of cancellations.

7.4.2 *Estimating cancellation type*

The next step of our framework predicts whether we expect a cancellation as a result of ticketing limits, fare movement, or departure date. Close to departure, cancellations are a result of time running out. For this reason, we refer to these as *"late" cancellations*. The majority of cancellations in between these time periods were driven by fare movements, which we allude to as **"middle" cancellations**. **"Early" cancellations** are cancellations that are driven by the ticketing limit running out.

Figure 7.5 shows the confusion matrix of estimating θ.

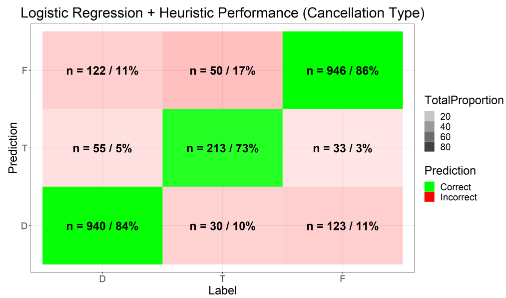


Figure 7.5: Confusion matrix for type of cancellation: D (departure date), T (ticketing) and F (fare movement)

Note that the predictions in Figure 7.5 are based on the heuristic of Equation (7.2). We first note that the model is most accurate in predicting early cancellations, followed by late and middle cancellations.

Let us now consider the "early" (ticketing process) predictions. 951 (86%) of 1107 early cancellations were classified correctly. For predictions that were incorrect, we note that the model is biased toward predicting a late cancellation, not a middle cancellation. For late cancellations, 946 out of 1123 observations were predicted correctly (84%). Again, there is a bias towards predicting an early cancellation. Note that this seems intuitively clear, since the number of middle cancellations are much lower than early cancellations. Finally, middle cancellations were predicted correctly in 73% (213 out of 293) of cases. For these cases, a bias toward early cancellations is present.

7.4.3 *Estimating the time processes*

In this section, we will study the performance of our estimates for the processes of ticketing (early), fare movement (middle) and time of departure (late).

7.4.3.1 *Ticketing*

Figure 7.6 shows the Beta (posterior) distribution of our belief of $P(X_d = 1 | X_{t-1} = X_{t-2}... = 0), d < t$ for different values of days d leading up to the ticketing deadline. Note that for $d = 5$, we have a flat (uniform) distribution. This is the result of only having a single observation in our dataset that cancels at $d = 5$. At $d = 4$ days before the ticketing deadline, we have observations which have cancelled, which caused our uninformative, uniform prior distribution to change into a density function with mass towards $p = 1$. At $d = 3$ we notice the biggest drop: the mean of this distribution centers about $p = 0.86$. Our estimates as d get proportionally smaller as we approach the ticketing deadline. Interestingly, the distribution of $d = 2$ has a similar variance than the distribution of $d = 3$, but is shifted to the left. The distribution of $d = 1$ exhibits an even smaller variability.

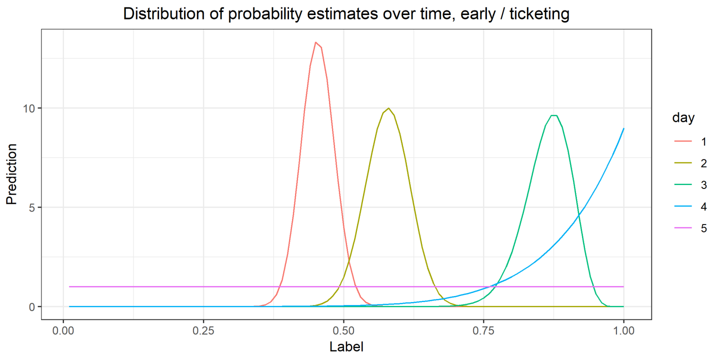


Figure 7.6: Probability distributions by day before the event. In this case, the days represent the number of days leading up to the ticketing deadline.

We have constructed a survival graph, in effect its cumulative distribution function. In Figure 7.14, in Section 7.4.4, we study this as a survival process.

7.4.3.2 *Time of departure*

Figure 7.6 shows the distributions of estimates for different days d.

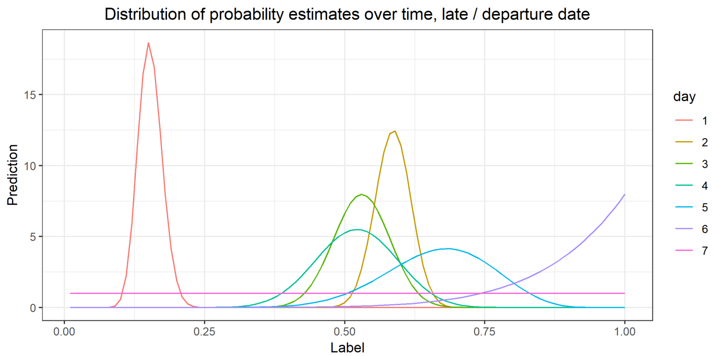


Figure 7.7: Probability distributions by day before the event. In this case, the days represent the number of days leading up to the departure date.

Note that in Figure 7.7, time moves backwards in time: day 1 represents one day before the day of departure. Note that this has a relatively low estimate, which peaks around $p = 0.14$. This estimate increases sharply on $d = 2$, to roughly $p = 0.59$. The largest number of cancellations happen on this day, which is shown again in Figure 7.14. Interestingly, the estimates for $d = 3$ and $d = 4$ are similar: $p = 0.57$ and $p = 0.58$, respectively. Moreover, the median estimate for $d = 5$ is still quite close at $p = 0.47$. From $d = 6$ onwards, the number of observations become small (namely, twenty or less) which explains the large variance in the distributions.

Figure 7.8 shows an example of the development of the distribution. In total, we have $85 + 76 = 161$ observations of this product that were alive before the start of $d = 2$ days before ticketing deadline. As discussed in Section 7.3.5, we start with an uninformative, Uniform(0,1) prior distribution. We then show how the posterior distribution changes once we obtain new observations. After $n = 18$ observations, our estimate is $p = 0.42$, but with a large variance. The colors represent a failure (non-cancellation) or success (cancellation) and show how the distribution changes accordingly. The final distribution is centered around $p = 0.52$.

Now, consider Figure 7.9. For this example, we used the same observations, with the only difference being our prior. For example, suppose we have indication that not every value of p is equally likely. We chose a $Beta(2,2)$ prior. Note how this changes our belief of p. At $n = 18$, our belief of p is 0.44. Also note that the distribution has a lower variability. As the number of observations grow, the importance of our initial belief, our prior, diminishes. We achieve the same distribution as we did with a $Uniform(0,1)$ / $Beta(1,1)$ prior.

Figure 7.8: Probability distributions by day after event, $d = 2$, $Uniform(0, 1)$ prior

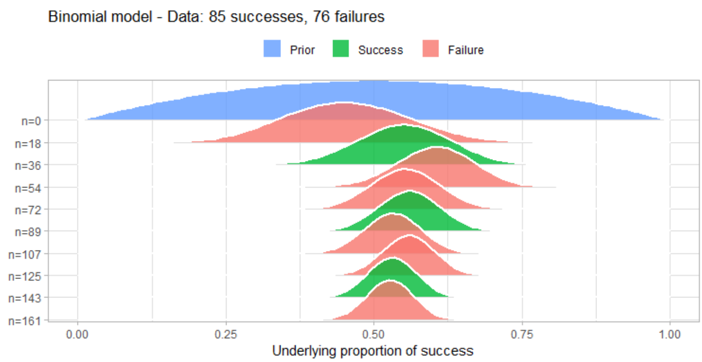


Figure 7.9: Probability distributions by day after event, $d = 2$, $Beta(2, 2)$ prior

However, because we propose a method that is dependent of both time of booking and booking class, in practice the number of data points may be very low. The importance of the choice of a prior distribution in this case, is shown in Figures 7.10 to 7.12. In this example, we took the first $n = 10$ observations from Figures 7.8 and 7.9, but used three different prior distributions: $Uniform(0, 1)$, $Beta(2, 2)$ and $Beta(4, 4)$ for Figures 7.10, 7.11 and 7.12, respectively. The choice of $Beta(4, 4)$ resembles the final (posterior) distribution best. This is particularly evident when looking at the distribution after a limited number of observations, say $n = 3$. When using $Uniform(0, 1)$ and $Beta(2, 2)$ prior distributions, there is substantial mass between $[0, 0.125]$ as compared to when using the $Beta(4, 4)$ distribution. Looking at the final posterior distribution in this case, after $n = 10$ observations, we note the final median estimates of $p = 0.57$, $p = 0.58$ and $p = 0.57$ for the aforementioned different prior distributions. However, its variance is lower when we used the $Beta(4, 4)$ distribution, which means we have obtained a more reliable estimate for our unknown parameter p.

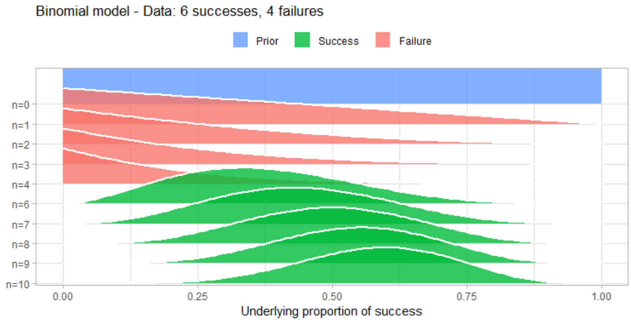


Figure 7.10: Probability distributions with a limited number of observations ($n = 10$), $d = 2$, $Uniform(0, 1)$ prior

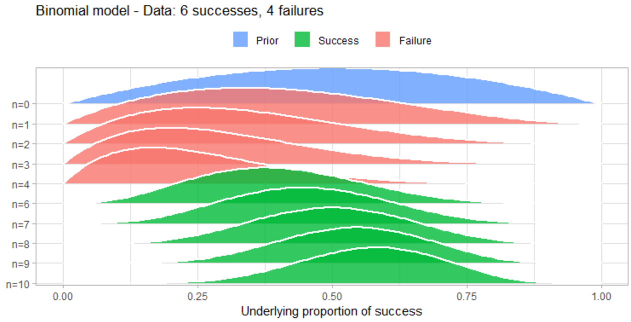


Figure 7.11: Probability distributions with a limited number of observations ($n = 10$), $d = 2$, $Beta(2, 2)$ prior

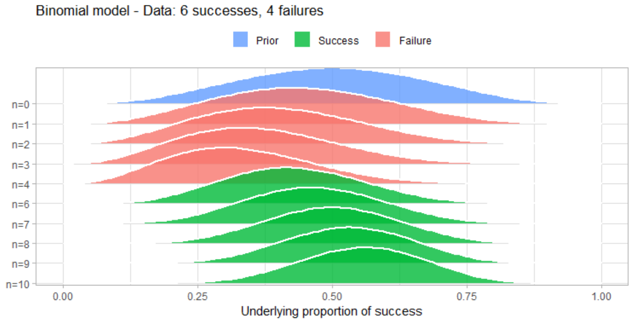


Figure 7.12: Probability distributions with a limited number of observations ($n = 10$), $d = 2$, $Beta(4, 4)$ prior

7.4.3.3 *Fare Movement*

Figure 7.13 shows the different distributions for values of d after a fare drop occurs. The distributions for $d = 0$ and $d = 1$ are not well defined, as the number of observations are not sufficient to build a reliable probability estimate, illustrated by the shape of the distribution. The distribution for $d = 2$ with a median estimate of $p = 0.1$ shows that 90% of bookings still alive do not cancel two days after a fare is lowered. These estimates grow substantially for $d = 3$ and 4, which exhibit similar estimates of $p = 0.52$ and $p = 0.54$, respectively. Moving to $d = 5$, this estimate grows again to $p = 0.68$. If a cancellation is affected by this process, the median estimate at $d = 6$ of $p = 0.86$ ensures that the vast majority of bookings still alive will be cancelled then. We also note that as d grows further, its variance grows.

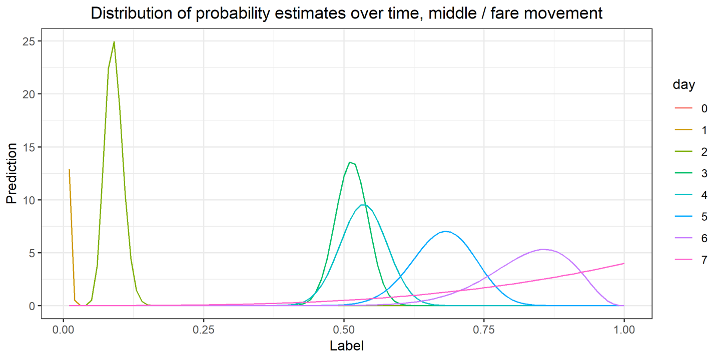


Figure 7.13: Probability distributions by day after event

7.4.4 *Comparison of processes*

In this section, we will first compare the distributions of parameters between processes to get an understanding of customer behavior.

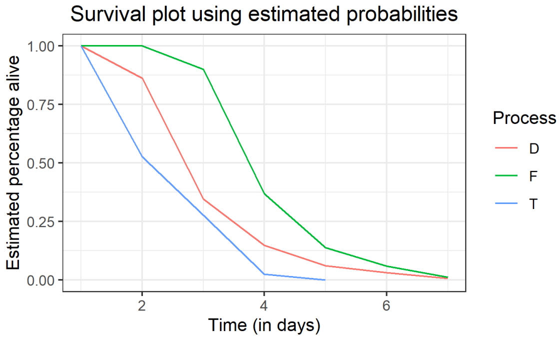


Figure 7.14: Survival plot using estimated probabilities for cancellation types cancellation D (departure date), T (ticketing) and F (fare movement). For the early (T) and late (D) process, the days represent the days leading up to the event. For the middle (F) process, the days indicate the time since the event happened.

Figure 7.14 shows a survival plot for the different processes: early, late and middle. This figure is generated by taking the median from Figures 7.6, 7.7 and 7.13 and calculating $P(X_t(\tau) = 1)$ as described in Section 7.3.6.

We make a few observations. First, the time to react in the middle process takes longer than early or late processes. Intuitively, this makes sense: for both the early and late processes, there are hard limits. A ticketing deadline or departure date are fixed. In theory, a customer may take weeks to react to a fare movement (even though in our work, we assume that decisions that are made one week after an event were not dependent on that event). Second, the probability of survival drops faster in the early process compared to the late process. For the early process, the survival rate seems linear between $d = 2$ and 4, for the late process, it seems to decay linearly between $d = 3$ and 5. Thirdly, the survival rate ends prematurely at $d = 5$, while it still has 11% and 27% of bookings alive for the late and middle processes, respectively.

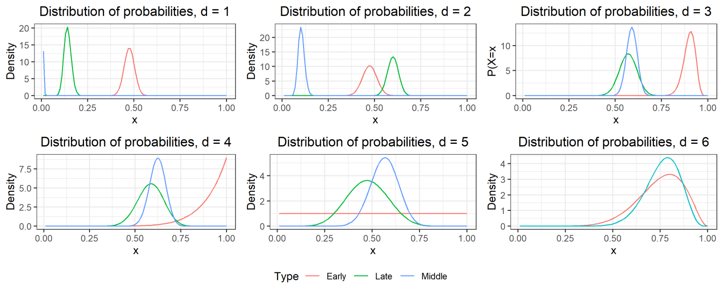


Figure 7.15: Comparison of distributions of parameter estimates for values of d

Figure 7.15 compares the parameter estimates for the three different processes by day d. These distributions are as shown in Figures 7.6, 7.7 and 7.13, but are easier to compare when fixing d.

Looking at $d = 1$, and comparing early with late processes, we observe that customers are much more likely to wait until the last day of ticketing than they are until the last day of departure. Very few customers react within a day of a fare drop.

This process is reversed at $d = 2$: here, a customer is more likely to cancel two days before departure than ticketing for the same d. Comparing the distributions at $d = 1$ and $d = 2$, we note that the parameter estimation is more accurate, illustrated by a distribution that is less spread out. At $d = 3$, the median estimates for the late and middle processes are similar, but the distribution for the late process is more spread out. This trend follows as d grows: the limits of the distributions grow and our estimates become less reliable. This is a direct result of a lower number of observations available at these values of d. Interestingly, the distributions of early and middle processes are very similar at $d = 6$.

7.4.5 *Estimating cancellation time*

In this section, we will discuss the results of the last step of our model: the actual time of cancellation. These results vary from OD to OD, from product to product, as every OD and product is given a separate model. In the results below, we will look at an average OD. For this OD, we were given a dataset with $n = 20946$ records. We obtained:

- an average error of -1.03 days,
- the mean absolute error of 6.5 days,
- for correctly classified types, this mean absolute error drops to 1.9 days.

Looking at the average error of 1.03 days, we conclude that our model, on average, as a whole, expects a cancellation earlier than they actually occur. However, the average mean absolute error is 6.5 days. This indicates that the prediction of cancellation is later than in reality. Classifying the type correctly, "early", "middle", or "late", this error drops to 1.9 days. This relationship shows the sensitivity and dependency on getting the prediction for cancellation type, Θ, right.

Table 7.6 shows the classification of early and late. The mean absolute error for late classifications, driven by departure date, is **1.49** days with a median of **1** day. For early classifications, driven by ticketing, these numbers are **1.72** and **2** for mean and median, respectively. We show that our model performs better at predicting "late" types. This is important since the exposure close to departure is much larger than earlier, at time of ticketing.

Type	Mean	Median
Early	1.72	2
Late	1.49	1

Table 7.6: Overview of classification and type and associated absolute mean error, assuming C_j and Θ are correct

We summarize our results in Figure 7.16.

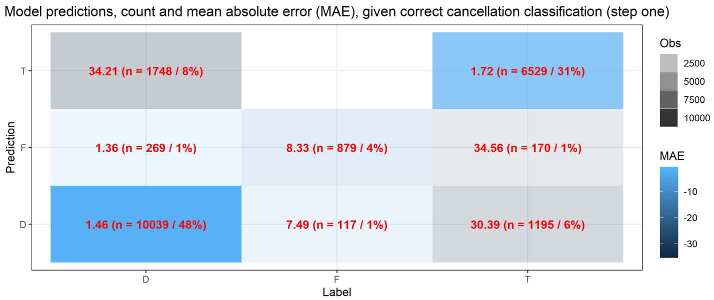


Figure 7.16: Overview of classification and type and associated absolute mean error, assuming the classification of C_j is correct for type of cancellation D (departure date), T (ticketing) and F (fare movement), $n = 20946$

From Figure 7.16 we can derive both the correct prediction of type of cancellation Θ and cancellation time estimate τ, as well as the distribution of early, late and middle cancellations and classifications. First, we note a total of 20946 observations, and note the distribution of types: 12056 of which were cancelled late (58%), 7894 early (38%), and 996 (roughly 4%) in between.

Out of 7741 early cancellations, we classified 6529 correctly. This amounts to 82.6%. If classified correctly, our mean absolute error (MAE) is 1.72 days. This error (expectedly) increases greatly to 30.39 days if we predict early as late, and 34.56 days if we predict early as middle.

Now consider the 12056 cancellations that were cancelled late. We did this correctly in 10039 cases, which represents 83.2%. In this case, the MAE is 1.46 days. Interestingly, if we predict a late case as middle, we obtain an even lower error at 1.36 days. We suspect that this may be the result of chance, because predicting late as early causes an MAE of 34.21 days, similar to the opposite error for early cancellations.

For the middle cases, we are able to predict 88.3% of the cases correctly, or 879 out of 996 cancellations. The model's performance for a correct type of prediction is worst here: 8.33 days. This seems to indicate that customer's response to a fare movement has a higher variance than the events of ticketing or departure date.

7.5 DISCUSSION

In this chapter, we introduced a framework to predict time of cancellation for airline tickets. This is a three-step process: first, we predict the probability that a given PNR cancels. Second, we predict what type of cancellation will occur: driven by ticketing, fare movement, or departure time. Third, given the type of cancellation, we estimate cancellation probabilities for different days before departure. One may wonder why we chose to approach this as a three step process.

The first step is clear: if a booking does not cancel, it does not have an associated time of cancellation. We therefore introduced two different models (a logistic regression and a gradient boosting model) to make this prediction. Obtaining the best model to predict C_j is outside the scope of our work - there may be other methods in the literature that perform better. Our framework is designed in such a way that it takes a 0/1-input for cancellation, and therefore flexible to be used with other methods.

The second step was a result of having identified three different, distinct processes. Through data analysis, see Figure 7.1, we show different behaviors based on time of booking and different processes. It became obvious that the majority of cancellations happen either shortly after booking, over time as a result of a fare movement, or shortly before departure. Each of these type of cancellations exhibit different response times. We also found a one-to-one relationship between time and type: we found a relation between time and process type: early/ticketing, late/departure date and middle/fare movement.

The last step in our framework concerns estimating probabilities by day. To estimate these probabilities, we used Bayesian Inference. In our work, we used an uninformative prior. It could be argued that certain features of a booking may be used to obtain an informative prior for different $p_{t,j}(t_i)$, introduced in (7.8). Suppose $p_{t,j}(1)$ and $p_{t,j}(1)$ are the probability estimates for cancellation one and two day(s) after a fare movement, respectively. For example, customers reacting the same day to a fare movement is less likely than say, two days after a fare movement. Intuitively, it makes sense there may be a time delay. We therefore could have assigned a prior with more mass towards 0 for $p_{t,j}(1)$, compared to $p_{t,j}(2)$. Especially in models that we created here, that are functions of both booking time t_b and booking class, where data is limited, an informative prior may improve model performance.

Comparing our results to our work on bookings, Chapter 6, which considered five of the same processes, it is interesting to note that only the process that describes the process of the airline's fare helped us model behavior. Interestingly, the alternative offer process, for example, while significant for bookings, is no longer a significant measure when estimating cancellations. This seems to indicate that once a customer has decided amongst a set of offers, it no longer considers other options, except for those driven by price, which is shown by the importance of the *Fal* process.

One difficulty in predicting cancellations is that there cannot be a cancellation without a booking was made first. Naturally, the number of cancellations cannot exceed the number of bookings, but if cancellations and bookings are forecasted independently, this may very well happen. In practice, this is typically solved by expressing the number of cancellations as a proportion of gross demand.

While this approach worked for the ODs we introduced in this chapter, it is not clear whether this approach will work for other airlines. If it does not seem to work, it would be interesting to study why this model doesn't work, and whether different underlying market characteristics are the cause of this. We see this as a future area of research.

The performance of our framework depends on the data quality: there may be OD's that have less data points than shown in Section 7.4. In cases where a limited number of data points is available, one may aggregate data points to POS/OD, similar POS/OD estimate C_j before estimating type. In these cases with a limited number of data points, we have stressed the importance of an informative prior to determine cancellation times. We have assumed that we have estimates for the time when an event occurs. Since we used the events of ticketing, flight departure, and fare movement, we can indeed reasonably estimate those. In fact, the time of ticketing and flight departure is fixed. And, as we discussed above, most airlines have a periodic fare review process which could be used in this model. Other processes, such as how the fare in the market moves, could also be introduced in this model. While this is a difficult task, as shown by Fiig et al [182], it is possible to predict one's own pricing behavior as well as competitor pricing within 10% mean absolute percentage error (MAPE).

In Section 7.4, we have conditioned on getting the estimate of cancellation correct, that is, we conditioned on accurately predicting C_j. After all, if make a mistake of type false negative - we predict a cancellation but this PNR will not cancel - we will have obtained a value for a most likely cancellation time, but we do not have a measure to compare this number to. This may call for a new metric to measure our work.

7.6 CONCLUSION

In summary, we have the following observations and conclusions:

1. Bookings that are made early are more likely to cancel sooner, rather than later.
2. Directly after booking, cancellations are predominantly described by whether it is ticketed.
3. Over time, bookings are cancelled because of a change in fare.
4. Closer to departure, time running out to cancel cause a rise in cancellations.
5. Three different processes can be used to determine time of cancellation at different times in the booking curve.
6. The ability of classifying whether an itinerary is cancelled (yes or no, regardless of time of cancellation) can be achieved through logistic regression.
7. Determining the type of cancellation may be achieved through a heuristic of three different logistic regression models.
8. Modelling underlying processes that affect cancellations over time can be done with a modified geometric distribution.
9. Estimating probabilities used in this geometric distribution can be achieved through Bayesian inference.
10. Misclassifying the type of cancellation affects the prediction of time of cancellation almost symmetrically, both causing the MAE to exceed of 34 days.
11. Upon correct classification, this framework produces forecast errors with MAE less than two days for ticketing and departure date processes, and 8.3 days for cancellations driven by fare movement.

8 OPTIMISATION WITH DOWNSELL AND DELAYED DECISION MAKING

Abstract

In this chapter, we study the impact of downsell in leg RM. Downsell happens when a customer purchases a lower fare than she was looking for. We aim to minimize the losses in revenue that arise from this situation. We reformulate the traditional DP formulation to account for this behavior by adjusting the input fare, and show significant revenue gains compared to the traditional DP formulation. Next, we aim to improve customer booking simulation by assuming customers may postpone their decision to book. Using a surprisingly easy reformulation of our DP strategy we ensure that cheaper classes will never open after they get closed, guaranteeing customers booking now is better than doing so in the future. Our results also show that when more than one eighth of passengers postpone their bookings, revenue gains are reported.

This chapter is based on [22].

8.1 INTRODUCTION

Over the past few years, there has been a major shift from the initial, simplified assumptions of RM, for example, from leg level optimization to network optimization. There has also been a big emphasis on customer choice and the corresponding willingness to pay, moving away from the classical independent demand assumption: demand between fare classes is assumed to be independent. It is clear that the independent demand assumption is invalid in practice. Especially with the recent trend of airlines removing fences that traditionally were able to segment customers (and thus create independency) relatively effectively.

However, most of the work is done on estimating sell-up probabilities. The goal here is to use these sell-up probabilities in the optimization problem to create a more realistic formulation. Estimating sell-up probabilities is not an easy task and requires a significant number of bookings in a network RM environment. While upsell can be seen as an "opportunity", one needs to be wary of the opposite:

down-sell, which can be seen as a "threat" in the formulation of typical RM problems.

Let us introduce our notion of downsell and upsell. When shopping for flights, we assume our customer to have a specific maximum price: say b. Using this definition, there is no notion of upsell: a passenger has a budget b and will only buy fares less than or equal to b, never more. Downsell occurs when a customer arrives, but finds a lower class (with or without the same conditions) available and books that class instead. If a fare lower than b is available, then this customer will buy this fare instead. In undifferentiated fare structures, every rational thinking customer will always book the lowest available fare. Approaching a customer's willingness to pay from this angle, there is no such thing as upsell.

In practice, the impact of downsell is most profound during the last stages of the booking window. We often observe a sharp drop in bid price close to departure, which opens up availability for the lower classes. While theoretically correct in this framework, this means that late arriving passengers, often with a high willingness to pay, buy the cheapest available fare. Therefore, it is very important to ensure the right availability for the right customer. Of course, this also calls for the airline's ability to segment their customers in such a way that minimal downsell occurs.

Another danger is present. A customer may decide to wait with her purchase, hoping that a lower class will become available. In this instance, a customer may have been willing to purchase a higher fare, but by waiting she takes advantage of a lower fare and downsell this way.

In fact, it is worth noting that airlines in the past have committed to reimbursing passengers when this happens. The so-called "guaranteed fare rule" meant that airlines would refund a customer the fare difference if their fare, with the same conditions, such as advance purchase, minimum and maximum stay requirements, dropped closer to departure. An example of this can be found in a news article published in the Sunday Deseret News [183]. Whether airlines these days still commit to this rule is unclear. For at least one major airline based in Canada, we have found evidence that this rule still applies.

In this chapter, we give insight in this problem and propose an alternative dynamic programming formulation which explicitly accounts for downsell. In Section 8.2 the problem at hand is formulated. This is followed by the methodology in Section 8.3. We assume that each customer will always buy the lowest available class, that is, there are no fare fences. The performance of this method is reviewed in Section 8.4. We extend this formulation to ensure that the lowest available fare will never drop, which is studied in Section 8.5. This leads to a remarkably simple and elegant formulation, but one which is highly relevant in practice, its results are shown in Section 8.6. This is followed by the robustness of this model in Section 8.7, and a discussion in Section 8.8. We close with conclusions in Section 8.9.

8.2 PROBLEM FORMULATION

One key assumption in RM is that the willingness to pay goes up as we get closer to departure: people *have* to travel. Therefore, it should be ensured that there are no possibilities for customers to purchase low-priced fares.

The traditional DP (DPID) leg level optimization can be found in, for example, [65]. This technique assumes independent demand. In practice we see that this method tends to work well when the ratio between demand and capacity exceeds a factor of 1.3. When faced with lower demand factors, we often see that the bid price does not achieve its desired effect. To illustrate this, let's consider an example.

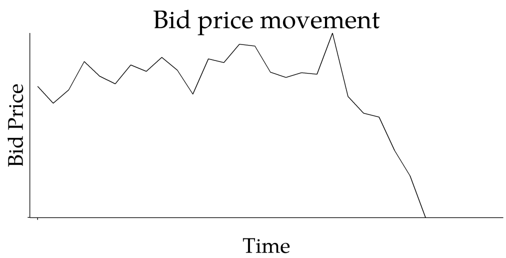


Figure 8.1: Example of bid price movement over time

Figure 8.1 shows how the bid price moves over the booking window for a long haul flight departure in a summer month. The demand factor is 1.02. This flight is forecasted to have about 10% of its demand for relatively high fares short before departure. In this Figure we see a sharp drop in bid price close to departure, despite forecasting demand in these later stages.

In the DPID method, we assume demand to be independent between classes: each arriving customer for a given class will actually buy the fare corresponding to that class. There is no notion of downsell. When the arrival process is different from the demand being used to calculate these bid prices, the bid price tends to drop and open availability to cheaper fare classes.

From a RM perspective, this means that the lowest filed fare is now available. This effect has two downsides. First, it causes arriving requests to downsell to the lowest fare throughout the booking window.

Second, if this also happens toward the stages close to departure, it does not incentivize people to book early. This may in turn cause customer behavior to change altogether, in which people will just "wait for a lower fare to become available".

This calls the need for a way to tell the optimization technique used to optimize revenue we expect downsell behavior. We will also investigate a way to avoid people from waiting for lower fares to become available.

8.3 METHODOLOGY

In this section, we show how downsell can explicitly be modelled in the dynamic programming formulation. We start with the DPID formulation found in [65]. Define:

λ_j is the arrival rate of class j, $j = 1, \ldots, J$;
f_j the fare of product/class j, $f_1 \geq f_2 \geq \ldots \geq f_J$
x the remaining capacity;
t the time unit, $t = 1, \ldots, T$.

Typically, demand for every product $j, j = 1, \ldots, J$ is expected to follow a Poisson process, with a mean λ_j. Time is then chosen in such a way that at most one arrival occurs. the fare may depend on t, but for simplicity we assume it is fixed over time t. Let use introduce the revenue-to-go function $V_t(x)$, which calculates the revenue to be earned having x seats and t units of time left until flight closure.

In order to formulate downsell, we first introduce the DP formulation which assumes independent demand:

$$V_t(x) = \sum_{j=1}^{J} \lambda_j(t) * max(f_j + V_t(x-1), V_{t+1}(x)) + (1 - \sum_{j=1}^{J} \lambda_j(t) V_t(x)) \tag{8.1}$$

The boundary conditions are:

$$V_{T+1}(s) = 0 \qquad s = 0, 1, \ldots, x \tag{8.2}$$

$$V_t(0) = 0 \qquad t = 1, \ldots, T \tag{8.3}$$

Equation (8.2) says that when all time has passed, the revenue-to-go is zero, regardless of the number of seats remaining: the flight has left. Equation (8.3) says that with zero seats left, regardless of the time unit, the revenue-to-go is zero. We solve (8.1) recursively backwards in time.

As Talluri and Van Ryzin ([65], p.59) point out, we check if $f_j \geq V_{t+1}(x) - V_{t+1}(x)$ and accept product j when this is true. This $V_{t+1}(x) - V_{t+1}(x-1)$ term is also known as the *bid price*, and provides an easy control mechanism. It provides monotonicity: when we consider two fares, $f_1 \geq f_2$, and when $f_2 \geq V_{t+1}(x) - V_{t+1}(x-1)$, we not only accept fare f_2, but also f_1.

In order to formulate downsell in the strategy of (8.1) we rewrite the value function as follows. Since we assumed that $f_1 \geq f_2 \geq .. \geq f_J$, this is equivalent to:

$$V_t(x) = \max_k \left(\sum_{j=1}^{k} \lambda_j(t) * (f_j + V_t(x-1)) + (1 - \sum_{j=1}^{k} \lambda_j(t)) V_t(x)) \right) \tag{8.4}$$

We call (8.4) the **DPID** formulation. It is easy to see why (8.4) is equivalent to (8.1): we look for the largest k, or similarly, the product with the smallest f_k which we are willing to accept. Since the fares are ordered, this automatically implies that we are also willing to accept classes $k-1, .., 1$. Thus, we have now turned this optimization problem into an optimization over classes k.

In the strategy of (8.4), it is now very intuitive to model downsell. Instead of obtaining revenue f_j, a customer will now buy the lowest fare available, that is, class k. The equation now becomes:

$$V_t(x) = \max_k \left(\sum_{j=1}^{k} \lambda_j(t) * (f_k + V_t(x-1)) + (1 - \sum_{j=1}^{k} \lambda_j(t)) V_t(x)) \right) \tag{8.5}$$

The approach in Equation (8.5) is referred to as the **DPDS** formulation. The optimal policies from Equations (8.4) and (8.5) is the value of k that maximizes the value function. These can be stored in a look up table for a given capacity and time units remaining.

Also note that every arriving customer now pays fare f_k. The amount of revenue lost by an arriving passenger j is therefore $f_j - f_k$. This is an interesting metric to investigate when we compare revenues generated by the DPID (8.4) and DPDS (8.5) policies by simulation.

As mentioned above, the control mechanism for (8.1) is easy to derive. A class j request is accepted whenever $f_j \geq V_t(x) - V_t(x-1)$. However, the control mechanism for strategy (8.4) is different. One cannot rely on the traditional notion of the bid price, $V_t(x) - V_t(x-1)$, anymore. The inequality $f_j \geq V_t(x) - V_t(x-1)$ no longer holds, since the fare that is used to calculate the value function may not be the actual fare of product j. Rather, it is the fare product that product

j will downsell to. Every passenger now pays exactly the bid price. Instead of a bid price table that has the bid price for a given x seats and t time units remaining, we construct a table that shows the class availability having x seats and t time units left. These are the values of k that maximize (8.5) for a given x and t. Instead of looking up a bid price, we use this look up table to decide whether to accept a class or not: if class k is open, we accept all classes $j \leq k$.

It is worth noting that in Equation (8.6) we find the efficient frontier (admittedly consisting of only one class) and solve the value function at the same time. In fact, this formulation is a special case of the value function in choice-set RM literature, such as [138], where the probabilities in the offered set are all zero, except for the lowest available fare (with probability 1). In effect, we calculate the efficient frontier for each combination of seats remaining x and time to-go t. This is a nice property as finding these efficient frontiers is computationally expensive when considering sets that have more than one option with a non-zero probability.

8.4 DPDS PERFORMANCE

In this section, we look at the model's performance with different demand levels. The data was adapted from real airline data, and is representative for a market where traffic is predominantly point-to-point. We compare the DPDS method with the DPID method and the EMSRb heuristic.

This data was then subsequently scaled to obtain high, medium and low demand factors. Table 8.1 shows the total demand for classes 1 through 5 for these demand factors, as well as the fares. The booking curves for these classes the medium case can be found in Figure 8.2.

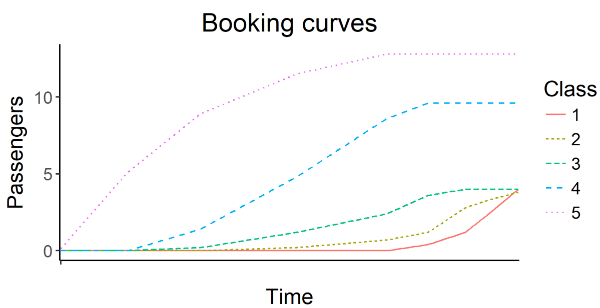


Figure 8.2: Booking curves

The booking curves for the high and low cases are similar but scaled to the numbers in Table 8.1. The fares are also given and are demand factor independent.

	Class demand					
DF	1	2	3	4	5	Sum
High	4	3.8	4	9.6	12.8	34.2
Med	3	2.8	3	7.1	9.1	25.0
Low	2.4	2.2	2.4	5.6	7.4	20.0
Fare	1800	1500	1000	800	400	

Table 8.1: Total expected demand and fares

We have performed simulations for each demand factor scenario and for each optimal policy for DPID, DPDS methods and the EMSRb heuristic. In what follows, we compare both revenues and load factors (LF) as KPIs. We have ran 100,00 simulations for each case to obtain a confidence level of 0.2%.

	DPID		DPDS		EMSRb	
DF	Rev	LF	Rev	LF	Rev	LF
High	13452	86.4	20481	89.0	17608	85.6
Medium	11007	86.0	16386	80.2	13437	85.2
Low	8239	75.8	13516	67.8	10161	74.7

Table 8.2: Performance of the DPDS method

From Table 8.2 we see that we experience load factor gains of over 2% in the high demand factor case, from 86.4% to 89.0%. This increase comes from selling to class 5 customers early in the booking cycle. In the high demand factor case, class 4 is shut early in the booking in cycle.

The standard deviation for load factors, not shown here, drops by almost one percent indicating slightly more robust load factors. While the load factor increased by 2%, the revenue gains amount to 54% compared to the DPID strategy in the high demand factor case. For comparison sake, we have included the performance using the EMSRb heuristic. The DPID method outperforms the EMSRb method as well. Note that the EMSRb method is more robust against downsell than the DPID method: the mean revenues are between 20% and 30% higher.

		Class				
DF	Method	1	2	3	4	5
High	DPID	0.1	0.6	0.7	8.9	11.5
High	DPDS	0.5	5.2	0.6	12.3	3.6
High	EMSRb	0.0	0.0	6.4	12.9	3.6
Med	DPID	0.0	0.4	0.5	4.3	16.4
Med	DPDS	0.1	4.4	0.1	9.2	6.4
Med	EMSRb	0.0	0.0	0.0	12.4	9.0
Low	DPID	0.0	0.1	0.1	0.9	17.9
Low	DPDS	0.0	3.5	0.0	7.5	5.8
Low	EMSRb	0.0	0.0	0.0	7.0	12.0

Table 8.3: Mean accepted passengers by class

Table 8.3 shows the average number of passengers sold by class. Here we see that the revenue gains come from selling less to class 5, and more to class 4 and 2. Table 8.4 shows that the average revenue lost as a result of downsell by demand factor. Here we see what the effects are of selling more to class 4 and class 2 passengers. The DPDS strategy outperforms both the DPID and EMSRb strategies.

Demand Factor	DPID	DPDS	EMSRb
High	-10088	-2377	-5628
Med	-8176	-2428	-4780
Low	-8314	-2095	-4054

Table 8.4: Mean lost revenue as a result of downsell

Looking at the medium demand factor case, we notice a drop in load factors, dropping by 5.8% from 86.0% to 80.2%. The average number of accepted passengers show that we protect against selling too much class 5. The majority of the passengers bought class 5, whereas with the DPDS strategy class 4 passengers are unable to downsell to this class and buy class 4 instead. The revenue we obtain using the DPDS strategy increases by 48.8% over the DPID strategy. The EMSRb performs better than the DPID strategy in revenue by about 20% and achieves a higher load factor. Compared to the DPDS method, however, we earn 25% less revenue. From Table 8.3 we see that in the medium factor case, the booking limits are not strict enough to avoid downsell: we only sell to class 4 and class 5. The controls by the DPDS methods lead to selling higher classes, particularly class 2.

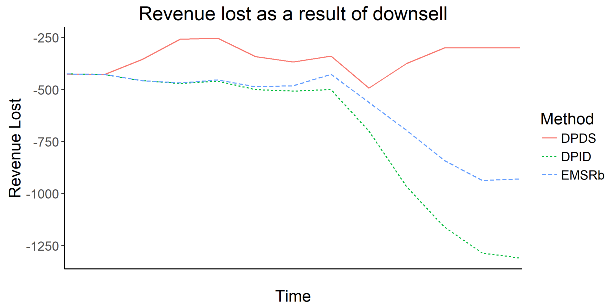


Figure 8.3: Lost revenue as a result of downsell

The biggest drop in load factor can be seen in the low demand case. Here, we sacrifice 8% and see the load factor drop from 75.8% to 67.8%. The revenue, however, increases by 64.6% from 8239 to 13561. It is to be expected that the biggest gains come from this scenario. The risk of downsell with low demand factors is inevitable, as shown in Table 8.3. While we only forecast 37% of our passengers to buy class 5, we end up selling 94% of our fares to class 5 customers, the direct effect of downsell. In the DPDS strategy, we only sell 34% of our flight to this lowest class.

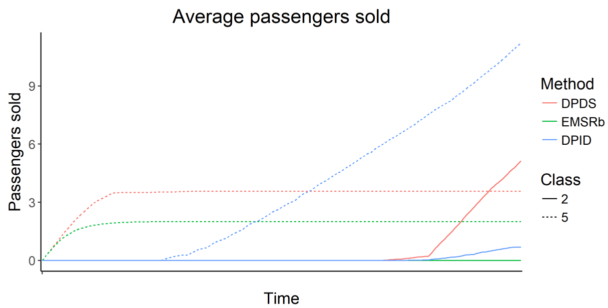
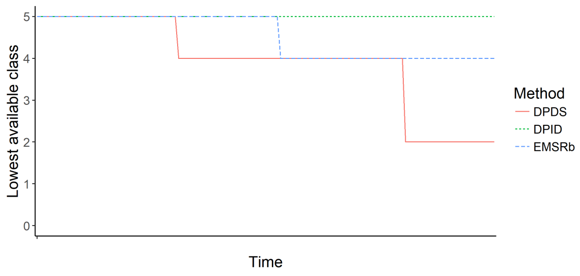


Figure 8.4: Lost revenue as a result of downsell

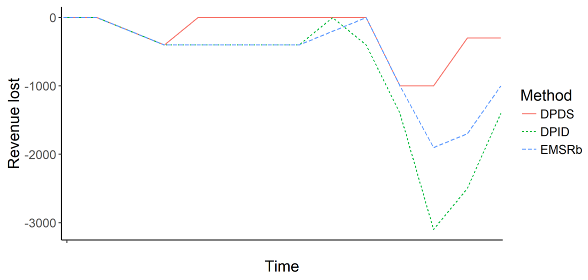
Figure 8.3 shows at what time revenue was lost because of downsell. This lost revenue is defined as follows. Let k be the lowest class available. An arrival j is accepted if $j \leq k$. The lost revenue is then $f_k - f_j$. These lost revenues are then bucketed by time units of 10 and averaged by simulation. The graph shows little differences in the early stages of the booking window: there is little difference in availability. Note that the DPDS does show lost revenue too, but it's controlled throughout the booking window. The DPID's lost revenue in the late stages can be explained by arrivals from class 1 which sell down to class 5 using this strategy.

Figure 8.4 shows the average number of passengers sold by time unit for the High demand factor cases for classes 2 and 5. We notice that in the early stages of the booking curve, the DPID strategy protects against class 5: it doesn't sell this class at all. It is too protective, in fact: later on it starts selling class 5. Both the EMSRb and DPDS strategies sell to class 5 early in the booking window, and curb their demand accordingly. Looking at class 2, the DPDS method is able to sell this booking class, with the EMSRb and DPID strategies selling a negligible amount.

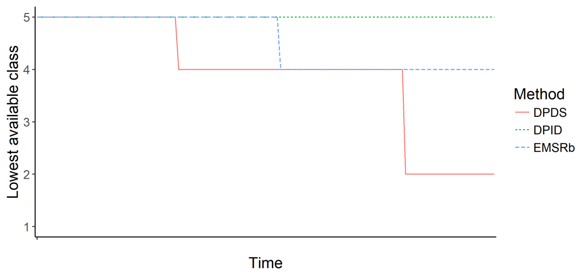
For the medium and low demand factor case (not pictured here) we see similar, but more extreme results: the DPID method is the only method which is able to actively sell to class 2 passengers. This corresponds to the revenue losses shown in Figure 8.3.



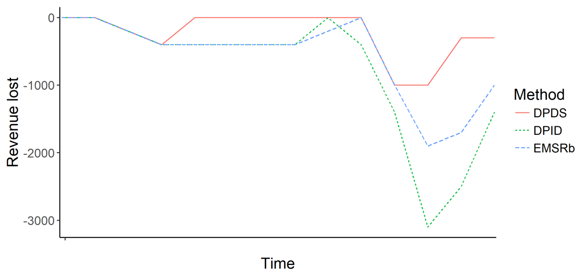
(a) Class availability - High demand factor



(b) Lost revenue because of downsell - High demand factor



(c) Class availability - Low demand factor



(d) Lost revenue because of downsell - Low demand factor

Figure 8.5: Single simulation of High (top) and Low (bottom) demand factor case.

Example

To illustrate where the revenue gains come from, let us consider two simulations from the high and low demand factor case. The first two charts in Figure 8.5 shows the high demand factor case.

Figures 8.5a and 8.5b shows the lowest available class for the DPID, DPDS and EMSRb methods on the left vertical axis for the high demand factor. For this simulation, the bid price only briefly reaches over 1000, which means class 3 is never closed - except for a few units in time. This can be seen from the lowest available class: class 4 is lowest available class for the majority of the booking curve. Close to departure, this goes up and down, before dropping back to class 5. The DPID availability close to departure shown here, determined by the bid price dropping, is similar to the one presented with real airline data in Figure 8.1.

The protection limit on class 5 with the EMSRb method causes class 5 to be closed earlier compared to the DPDS method. However, later in the booking window, when demand from high classes is expected, the EMSRb method closes class 4: the lowest available class is class 3. The DPDS methods closes classes gradually. This translates to the lost revenue in Figure 8.5b. This figure shows the total revenue lost over time. In the beginning, there is only demand from class 5, so neither method loses revenue as a result of downsell. The effects are most profound in the later stages: here we find the DPID method losing most revenue, followed by the EMSRb method. The DPDS method avoids substantially less lost revenue.

Figures 8.5c and 8.5d shows the same, but for the low demand factor case. The bid price never exceeds 400, which translates into the availability line plot for the DPID. The DPDS method closes class 4 before the EMSRb method, and toward the end closes down to class 2. On the right in this Figure, we see the effects of lost revenue as a result of downsell for the three different methods. Having class 5 open for the DPID method results in the most lost revenue, followed by EMSRb, which has class 4 open, followed by DPDS, which has class 2 open at the late stages.

8.5 WAITING BEHAVIOR - DPDS$\uparrow$

As we mentioned in Section 8.2, one of the dangers of a dropping bid price close to departure is, in absence of advance purchase restrictions, discouraging passengers from booking early. With the strategy given in equation (8.5) it could still happen that cheaper classes open close to departure. One way of avoiding this from happening is to ensure that the lowest class available won't drop over time for a given remaining capacity x.

We accomplish this by incorporating the current lowest class available in our state space. Up to this point, the state space only consists of the number of seats available: (8.1) and (8.5). Therefore, we introduce a new variable to our state space, y, that denotes the lowest class available:

$$V_{t,y}(x) = \max_{k:k \leq y} \left(\sum_{j=1}^{k} \lambda_j(t) * (f_k + V_{t+1}(x-1,k)) + (1 - \sum_{j=1}^{k} \lambda_j(t) V_{t+1}(x,k) \right) \tag{8.6}$$

We dub this optimization technique by DPDS↑, which indicates that the class availability will always be non-decreasing.

Note that now we no longer have the option to choose any k. We now have $k \leq y$, indicating that the choice of class k should be less than or equal to the current lowest available class y.

For every number of seats remaining x, we initialize the value of y at $t = T$ to 0. That is, at departure, the lowest class available is 0: there are no more seats to left. This is the extra boundary condition we introduce here.

8.6 DPDS↑ PERFORMANCE

In this section, we compare the relative performance of the DPDS↑ method with DPDS method from the previous section.

The simulation experiment was designed as follows. A customer enters the system wanting to purchase a given class k. His behavior is different for the DPDS and DPDS↑ optimization techniques. For optimization using DPDS, this customer will wait with a probability p. The waiting time for this customer is uniformly. The maximum number of retries is set to three. This same customer will not wait with purchasing in the DPDS↑ case, as we can guarantee that this customer the price will not drop. In both cases, a customer will buy the lowest available class.

The relative performance of the DPDS↑ for the medium demand factor is illustrated in Figure 8.6.

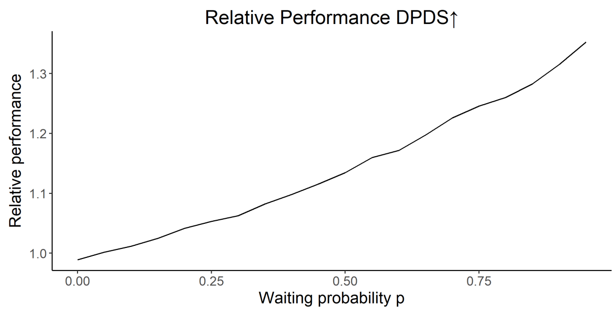


Figure 8.6: Relative performance of DPDS↑, high demand factor case

The performance is shown as a function of the waiting probability p. Note that $p = 0$ is a special case where customers never wait with their purchase. Observe that the relative revenue performance loss in this (downside) case is well-controlled, with a loss of -2.3%. This is a direct result of a more restrictive DP formulation: if a class is closed we may never open it again.

From $p = 0.12$ we see revenue gains. To illustrate where these revenues gains come from, consider Figure 8.7.

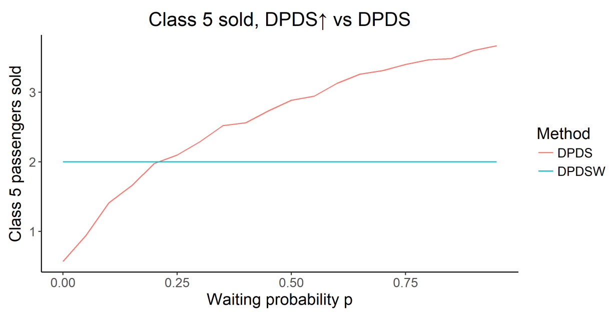


Figure 8.7: Passengers sold to lowest class, high demand factor

Figure 8.7 shows how many passengers buy the cheapest class, for the high demand factor case, as a function of the probability of postponing p. If nobody decides to wait, that is, $p = 0$, the DPDS↑ method performs worse and sell more of the lowest class. From around $p = 0.12$ we see the DPDS↑ sells less class 5. The more people wait, the more they profit from class 5 opening up with the DPDS strategy.

It should be noted that deep analysis of these results is outside the scope of this chapter. This only gives an impression of the revenue effects of delay behavior. The revenue effects are dependent on several parameters, including waiting probability p, the delay distribution and number of retries. We discuss this in more detail in the discussion section.

8.7 robustness

In this section, we discuss the robustness of the DPDS↑ method. There is an inherent downside of this method: once a class is closed, it may never be opened again. This means that if we overestimate demand, the controls may be too aggressive. Similarly, if we underestimate demand, the controls may be too lenient: it is rarely a good idea to close a class. Therefore, it is important to study the robustness of the DPDS↑ method with respect to under- and overforecasting demand.

Table 8.5 shows the robustness of the DPDS↑ method. The robustness of the DPDS method (with a moderate choice of $p = 0.25$) was added for sake of comparison. To analyze the results, we use the superscript to denote what demand factor is used to optimize, and the subscript to indicate what demand factor is realized (simulated). For example, V_L^H means having determined availability for the *H*igh demand case, but realized a *L*ow demand factor in the simulation.

Actual:	High		Medium		Low	
Optimized for:	DPDS	DPDS↑	DPDS	DPDS↑	DPDS	DPDS↑
High			0.74	0.78	0.59	0.64
Medium	1.32	1.15			0.77	0.80
Low	1.68	1.37	1.23	1.21		

Table 8.5: Robustness of DPDS and DPDS↑ methods compared

Table 8.5 shows that the DPDS↑ is similar in its robustness compared to the DPDS method when it comes to overforecasting (i.e., V_M^H, V_L^H and V_L^M). What is lost in load factor is made up in yield. It is less robust compared to the DPDS method in the underforecasting cases: V_M^L, but particularly V_H^L and V_H^M. In these cases, downsell is actually more prevalent than what we see in the DPDS strategy as the DPDS↑ strategy doesn't close classes quickly enough. Comparing the DPDS↑ with DPDS formulations, we can conclude that the former is more robust to overforecating, while the latter is more robust to underforecasting.

8.8 discussion

In this chapter, we have shown the dangers of the popular, traditional leg-level dynamic programming method, DPID. By changing the value function to explicitly account for downsell, we are much more able to sell the right fare to the right customer. This contributes to significant revenue improvements. Compared to the work by Fiig et al. [131], where fares are adjusted in the optimisation, we adjust the fare *before* the optimisation. We have also shown how to extend this strategy to ensure that once a class is closed, this class will never open again. As we mentioned, there may be business reasons for this. One example we gave is that having cheap classes available close to departure "doesn't look good".

It should be noted that while this chapter was focused on airline RM, the same methodology can be used by other industries. For example, hotel and car rental companies practice RM as well and they too run the risk of people canceling their reservation and rebooking should the purchased rate drop over time. For more reading on cancellations and their impact, we refer the reader to Sierag [144].

The controls produced in this chapter are leg-level controls: they show what class should be open or closed. However, it is not entirely clear how controls are used in a network. In the traditional DPID case, one accepts request if the fare exceeds the sum of the bid prices. It is not immediately clear how these controls would work in a network case, this could be pursued in future work.

As we have shown, the revenue differences between the DPID and DPDS strategies are profound. We stress that the results in Table 8.2 are highly dependent on the demand estimates, fares, and capacity of the flight. In our case, as we show in Table 8.1, the fare ratio between the most expensive and least expensive fare is 4.5. While for the data we have been provided that that his holds true, we realize that this may not be representative for all flights. Lower fare ratios will result in lower revenue gains over DPID. However, it remains a very relevant problem, regardless of fare ratios.

The results are also dependent on the demand factors. The greatest revenue gains can be realized in low demand factor situations. This is intuitively clear: with low demand, one can expect a lot of availability and thus the risk of downsell being most profound.

It is worth noting that using our simulation setup, DPID is consistently outperformed by EMSRb, regardless of demand level. This is something we have found in Table 8.3, where we observe that the EMSRb method is much better at protecting against class 5 than the DPID method. It seems that with complete downsell, EMSRb outperforms DPID. To understand these findings, we also simulated the effects without downsell; that is, every arriving class k buys fare f_k. Here, the DPID performed best, with EMSRb followed behind by 1.4%. The DPDS strategy performed 2.9% worse than the DPID strategy. This seems to confirm hypothesis: downsell enables EMSRb to outperform the DPID strategy.

We also showed that the gains in revenue come as a direct result of achieving a better class mix. In this chapter, we have assumed that everyone will buy down. However, it may not be true that a class 1 passenger will in fact buy down to class 5. If classes 1 through 5 are distinctively different in conditions, such as its advance purchase or minimum stay, downsell from the most expensive to the most cheapest class is not likely. However, this may be true for low cost airlines where only price points make products differ. Moreover, the current trend in airline pricing in general is removing fare fences, so the results we have show are representative. One can avoid this extreme buydown by keeping fences.

Most airlines remove fences between fares, but most of them do have fare families. A fare family is a set of classes with the same conditions, but different fares. It is interesting to reformulate equation (8.4), so that customers will only downsell to the lowest fare within the fare's fare family.

We want to stress the importance and implications of the DPDS↑ strategy. When an airline is able to communicate to its customers that the fare will never drop, there is very little incentive for customers to delay their purchase. Firstly, this has the potential to change customer behavior, and in turn, may stabilize bookings and improve our ability to forecast demand more accurately. Secondly, this prohibits people with booking an expensive, flexible fare early and rebooking to a cheaper fare once their travel plans have been confirmed and flexibility is no longer required.

In fact, airlines that operate flights out of the United States of America are already at risk of delayed-decision making. The Department of Transportation (DOT), the governing body that oversees among other industries, aviation, requires airlines to offer a full refund to customers if they cancel within 24 hours of booking, see [184]. One of the situations that may arise is that people do book a flight, find the fare drop within 24 hours, and cancel and book a new ticket. While this technically is not delayed-decision making, the customer made a purchase after all, this is something that can be avoided by using the DPDS↑ strategy.

Lastly, we have made some arbitrary choices modelling the "waiting" behavior: a uniform time between retries, and three retries in total. To validate these assumptions, an elaborate data analysis should be carried out. This falls outside the scope of this chapter and is part of our future research plans. This also calls for a concrete definition of what "waiting" really is.

8.9 CONCLUSION

In summary, we have the following observations and conclusions:

1. The traditional dynamic programming formulation (DPID) assumes independent demand: a customer either purchases a specific product, or it doesn't.
2. Having identified that price is frequently the most determining factor in the decision process, found in Section 2.2.1 and Chapter 6, we introduce a new formulation assuming a customer always purchases the lowest fare possible.
3. This new formulation is a surprisingly easy reformulation of the traditional formulation. Very promising results are shown, with revenue increases of up to 52%.
4. By means of website analytics data, we show that most customers search for an itinerary but do not make a purchase later; rather, they postpone their decision.
5. Over time, this may cause lower-priced classes to open up, hurting the airline but benefiting the customer. This frequently happens over time.
6. We introduce a new optimisation method that ensures that the fare offered by the airline is always non-decreasing.
7. Through a simulation study, we show that if more than one eighth of the customers postpone their decision, revenue gains are found.
8. This method does not only improve revenues, but also indirectly improves forecasting methods: a customer is no longer incentivized to postpone their decision, therefore reducing the variability in bookings and in turn, making forecasting easier.

9

OPTIMISATION USING CANCELLATIONS

Abstract

In this chapter, an optimisation method is introduced that accounts for cancellations. We account for the cancellation by estimating the opportunity cost of this booking between the time of booking and the expected time of cancellation. The probabilities that are required are found using the framework of Chapter 7. The dynamic programming formulation we introduce remains single-dimensional, which is important for this algorithm to be implemented in practice. The formulation involves an estimate of the value of the state of the system at the time of cancellation (which is in the future), which is found by making a choice of novel heuristics which we introduce. The fare that is used to determine whether a product is available for sale, is adjusted by the risk the airline faces. We introduce an example which shows that there may be cases where it is optimal to reject a higher-priced product if the risk of cancellation is high, while accepting a lower-priced product. Through simulation studies, we show increases in revenues against a traditional dynamic programming formulation that does not explicitly models cancellations. Next, we show that the optimisation method is robust against choice of heuristic, misjudgement of cancellation probability and forecasting errors.

This chapter is based on [23].

9.1 INTRODUCTION

Cancellations in RM pose a risk to the company trying to maximize revenue. There are two different risks. The first risk, *is not being able to resell the unit of capacity (a hotel room, rental car, seat on a plane) before the product offering perishes (when time has passed, such as a night has passed, or a plane has taken off).* The second risk, is an implicit risk: *bookings on hand limit the number of units of capacity for sale. An increase in bookings on hand typically imply a positive price change, which in turn means a smaller customer base.* In the remainder of this chapter, we will focus on the these problems in the context of the airline RM problem.

In the airline RM, cancellation rates vary greatly by point-of-sale (POS). The POS represents in what country a ticket is sold. Cancellation policies often depend on both the POS and origin and destination (OD) pair, but behavior is typically driven by POS. Cancellation policies are therefore set accordingly by POS. Consequently, cancellations vary greatly by POS: naturally, POS's with low cancellation fees demonstrate a higher percentage of cancellations than POS with high fees. In practice, cancellation rates range from 20% to 60% in some cases.

The aforementioned risk of cancellation depends on time, as well as on the number of expected cancellations. Consider the time aspect. A booking and a corresponding cancellation one year before departure has a very low risk of type the first type of risk: there is sufficient time to resell this seat. Similarly, second type of risk is low since early on in the booking curve, capacity is likely still low and therefore plenty of availability. However, now consider a booking a year before departure and a cancellation an hour before departure. In this case, there is a total risk of the first type: there is no time left to resell this unit of capacity. The risk of type two depends how full the flight is: if it is at or near capacity, this one seat will have had an effect on availability and others would have been unable to purchase.

In practice, airlines combat this problem by overbooking. Overbooking is the process of selling more seats than physical capacity. Overbooking too many seats has financial consequences to the airline: in European territories, an airline is obliged to pay up to €600 and reaccommodate a passenger at the earliest possibility, even if this means rebooking on other airlines. Outside European territories, only the latter applies, but this still comes at a cost. Similarly, overbooking not enough results in empty seats which could have been sold. This shows the need for an accurate cancellation forecast and corresponding optimization.

In our work, cancellation rates do not only depend on time, but also on time booked. We have found substantial evidence that rates depending on class are not sufficient, but, rather, the time of booking is just as, if not more important. This is something that was shown in [21].

This chapter is structured as follows In Section 9.2, the formulation without cancellations is introduced. In Section 9.3 we introduce our dynamic program with cancellations. Section 9.4 covers results where we compare these two formulations. A discussion is given in Section 9.5. We provide conclusions in Section 9.6.

9.2 TRADITIONAL FORMULATION (NO CANCELLATIONS)

We begin by repeating the traditional dynamic programming formulation, introduced by Talluri and Van Ryzin [65]. Consider:

λ_j is the arrival rate of class j, $j = 1, \ldots, J$;
f_j the fare of product/class j, $f_1 \geq f_2 \geq \ldots \geq f_J$;
x the remaining capacity;
t the time unit, $t = 1, \ldots, T$.
$R(t)$ a random variable, with $R(t) = p_j$ if a demand for class j arrives, and 0 otherwise.

Suppose we have discretized time in such a way that in each time slice, we can have at most one arrival. Also note that $P(R(t) = p_j) = \lambda_j(t)$. When presented with an arrival, we need to decide whether to receive the current revenue, given by the random variable R_t, and move to the next time unit with one unit of capacity less, or reject this arrival request but have the same number of capacity in the next time unit. Therefore, introduce an indicator variable $u \in (0,1)$, which is what we want to maximize over:

$R(t)u + V_{t+1}(x-1)$

Now define $V_t(x)$ as the value function that represents the expected revenue-to-go given t units of time left and x units of capacity.

$$V_t(x) = E\left(\max_{u \in (0,1)} R(t)u + V_{t+1}(x-u) \right) \tag{9.1}$$

We denote Equation (9.1) as **DPID** to indicate this is the traditional dynamic programming formulation, without cancellations.

Equation (9.1) implies that $u = 1$, that is, accept a given request, if and only if:

$$\begin{aligned} R(t) * 1 + V_{t+1}(x-1) &\geq R(t) * 0 + V_{t+1}(x) \\ \to R(t) + V_{t+1}(x-1) &\geq V_{t+1}(x) \\ \to R(t) &\geq V_{t+1}(x) - V_{t+1}(x-1) \end{aligned} \tag{9.2}$$

Having identified this, introduce:

$$\Delta V_t(x) = V_t(x) - V_t(x-1). \tag{9.3}$$

$\Delta V_t(x)$ can be thought of a point-estimate for the gradient in the x direction. Now, taking the expected value of $R(t)$, we obtain the optimal policy:

$$\begin{aligned} &\text{Accept a given product } j \text{ if and only if:} \\ &f_j > \Delta V_{t+1}(x) \end{aligned} \tag{9.4}$$

9.3 NEW FORMULATION

Define f_j as the fare of product j. The refund percentage is given by ρ_j. Therefore, the value returned to the customer if cancelled is equal to $\rho_j f_j$. Let $c_j(t,\tau)$ be the probability that product j booked at time t is cancelled at time τ. For more information how this is calculated, please refer to Chapter 7. Let $\zeta_j = \sum_{\tau=t+1}^{T} c_j(t,\tau)$, such that $1-\zeta_j$ represents the probability that product j is not cancelled.

To define our new formulation, we consider two new random variables. $R(t)$, is as before, represents the direct reward for accepting product j. The second random variable, $C(t)$, represents the future (negative) revenue of the system of a cancellation. Next, we introduce $O(t,x(t))$ which is the opportunity cost of one seat up to time t. This opportunity cost is of course dependent on the state of the system x at time t.

$$R(t) = \begin{cases} f_j + V_t(x-1) & \text{if a request arrives for product } j \\ 0 & \text{otherwise} \end{cases} \tag{9.5}$$

Since a request for product j arrives using a Poisson process with $\lambda_j(t)$, we have $P(R(t) = f_j + V_t(x-1)) = \lambda_j(t)$.

$$C(t) = \begin{cases} -\rho_j f_j & \text{if a request for cancellation for product } j \text{ occurs} \\ 0 & \text{otherwise} \end{cases} \tag{9.6}$$

Equation (9.6) says that the future reward depends on whether a cancellation request arrives. If a product cancels, the airline has to refund $\rho_j f_j$ to the customer. If it does not cancel, there is no (negative) revenue. A cancellation request having booked at time t occurs at time τ with $c_j(t,\tau)$, so we have $P(C(t) = -\rho f_j) = c_j(t,\tau)$.

Next, introduce:

$$O(\tau, x(\tau)) = \begin{cases} V_\tau(\tilde{s}) - V_\tau(\tilde{s}-1) & \text{if a product } j \text{ is cancelled} \\ 0 & \text{otherwise} \end{cases} \tag{9.7}$$

Similarly, we only suffer an opportunity cost in the future if a booking will cancel (if it doesn't, it is only evaluated against the opportunity cost $V_{t+1}(x) - V_{t+1}(x-1)$), so we have $P\big(O(\tau,x(\tau)) = V_\tau(\tilde{s}) - V_\tau(\tilde{s}-1)\big) = c_j(t,\tau)$. It is not cancelled, we do not incur any future opportunity cost. Of course, a-priori we do not know this, and we also do not know what state, $\tilde{s}$, we are in at a future time τ.

Equation (9.7) says that if a product is cancelled at time τ, with probability $c_j(t,\tau)$, the airline lost the opportunity of that one seat up to that point τ. We use the $\tilde{s}$ to indicate that this itself is a random variable. This term, $V_\tau(\tilde{s})$ term itself introduces complexity. After all, since a-priori we don't know when this request when this request will come. Similarly, we do not know what state (how many seats) the system will be in. In the next section, we will discuss simple heuristics to estimate $V_\tau(\tilde{s})$.

The expected reward from accepting a booking is then equal to:

$$E[R(t)] + E[C(t)] - E[O(t)] \tag{9.8}$$

We make this distinction so that it is clear to the reader what the revenue is that the airline expects: it receives a direct revenue given by $R(t)$, has to offer a refund given by $C(t)$ if cancelled, and loses $O(t)$ revenue in between. Therefore, the new Bellman equation follows that of Equation (9.1), and is equal to:

$$V_t(x) = E\left(\max_{u \in (0,1)} u\big(R(t) + C(t) - O(t)\big) + (1-u)V_{t+1}(x) \right) \quad \textbf{(DPC)} \tag{9.9}$$

Similar to the naming convention of Equation (9.1), we name Equation (9.9) as **DPC**, our dynamic programming formulation including cancellations.

The optimal policy can be derived similarly. Maximizing the term, taking $u = 1$, or, similarly, accepting a product j, happens if and only if:

$$\begin{aligned} &1 * \Big(R(t) + C(t) + O(t)\Big) + (1-1)V_{t+1}(x) > \\ &0 * \Big(R(t) + C(t) + O(t)\Big) + (1-0)V_{t+1}(x) \\ &\rightarrow E[R(t) + E[C(t)] + E[O(t)] > V_{t+1}(x) \end{aligned} \tag{9.10}$$

We will derive the optimal policy below.

$E[R(t)] + E[C(t)] + E[O(t)] > V_{t+1}(x)$, which implies

$$=f_j + V_t(x) + \sum_{\tau=t+1}^{T} c_j(t,\tau)\Big(-\rho_j f_j\Big) + \Big(1 - \sum_{\tau=t+1}^{T} c_j(t,\tau)\Big) * 0$$

$$-\sum_{\tau=t+1}^{T} c_j(t,\tau)\Big(V_\tau(\tilde{s}) - V_\tau(\tilde{s}-1)\Big) + \Big(1 - \sum_{\tau=t+1}^{T} c_j(t,\tau)\Big) * 0 > V_{t+1}(x)$$

Removing zero-valued terms, we obtain:

$$=f_j + V_t(x) + \sum_{\tau=t+1}^{T} c_j(t,\tau)\Big(-\rho_j f_j\Big) - \sum_{\tau=t+1}^{T} c_j(t,\tau)\Big(V_\tau(\tilde{s}) - V_\tau(\tilde{s}-1)\Big) > V_{t+1}(x)$$

Since $-\rho_j f_j$ does not depend on τ, and we have defined $\zeta_j = \sum_{\tau=t+1}^{T} c_j(t,\tau)$, we have:

$$=f_j + V_t(x-1) - \zeta_j \rho_j f_j - \sum_{\tau=t+1}^{T} c_j(t,\tau)\Big(V_\tau(\tilde{s}) - V_\tau(\tilde{s}-1)\Big) > V_{t+1}(x)$$

collecting terms, we obtain:

$$=f_j(1-\zeta_j\rho_j) + V_t(x-1) - \sum_{\tau=t+1}^{T} c_j(t,\tau)\Big(V_\tau(\tilde{s}) - V_\tau(\tilde{s}-1)\Big) > V_{t+1}(x).$$

Finally, rearranging terms, we have:

$$=f_j(1-\zeta_j\rho_j) - \sum_{\tau=t+1}^{T} c_j(t,\tau)\Big(V_\tau(\tilde{s}) - V_\tau(\tilde{s}-1)\Big) > V_{t+1}(x) - V_{t+1}(x-1)$$

Note that the right hand side is as before, the opportunity cost of capacity in the next time unit.

Let us define $f_j^* = f_j(1-\zeta_j\rho_j)$. Next, define $\Delta V_t(x) = V_t(x) - V_t(x-1)$. Substituting these terms, we obtain the following policy, given in Equation (9.11):

Accept a given product j if and only if:

$$f_j^* > \Delta V_{t+1}(x) + \sum_{\tau=t+1}^{T} c_j(t,\tau)\Big(V_\tau(\tilde{s}) - V_\tau(\tilde{s}-1)\Big) \tag{9.11}$$

The term on the left, f_j^*, can be seen as the expected value of the fare of a request. It is given by the fare adjusted by its cancellation probability ζ_j, as well as its refund percentage ρ_j.

The first term on the right, $\Delta V_{t+1}(x)$, is the opportunity cost of a unit of capacity in the next time stage. Next, consider the $V_\tau(\tilde{s}) - V_{t+1}(x-1)$ term inside the summation. This difference shows an opportunity cost between time unit t and time of cancellation τ. This opportunity cost is weighted by the probability of time of cancellation τ. Equivalently, we can write Equation (9.11), as follows:

Accept a given product j if and only if:

$$f_j^* - \sum_{\tau=t+1}^{T} c_j(t,\tau)\Big(V_\tau(\tilde{s}) - V_\tau(\tilde{s}-1)\Big) > \Delta V_{t+1}(x) \tag{9.12}$$

This formulation and corresponding optimal policy of Equations (9.1) and (9.4), respectively, are a special case of Equations (9.9) and (9.11). After all, in absence of cancellations, we have $\zeta_j = 0$ and $\rho_j = 0$ for all j, so we have $f_j^* = f_j(1 - \zeta_j\rho_j) = f_j$. Similarly, no cancellations imply $c_j(t,\tau) = 0$ for all t, τ, j, so the summation of (9.11) disappears and reduced to $f_j > \Delta V_{t+1}(x)$, which was shown in Equation (9.4).

The problem is solved by substituting the optimal policy of Equation (9.11) into Equation (9.9):

$$\begin{aligned} V_t(x) &= E[R(t) + C(t) + O(t)] \\ &= \sum_{j=1}^{n} \lambda_j(t)\left(f_j^* - \sum_{\tau=t+1}^{T} c_j(t,\tau)\Big(V_\tau(\tilde{s}) - V_\tau(\tilde{s}-1)\Big)\right)^+ \end{aligned} \tag{9.13}$$

Where the $^+$-notation in (9.13) indicates that we take the maximum of this term and zero - therefore, the term is replaced by zero if negative (and, equivalently, if the left hand side of Equation (9.12) is not greater than the right hand side).

Equation (9.13) shows the power of this heuristic: the state-space is still one-dimensional, x. In practice, it is very important to make this a feasible approach. In Chapter 4, we discuss practical limitations of the RM system. Specifically, in Section 4.4.1 it is discussed that reoptimizing the most basic formulation, Equation (9.1), is impossible to do daily in practice. Therefore, the work of Sierag [144], for example, which uses more than a single dimension in the state space, cannot easily be used in practice and it is critical to keep the state space one-dimensional.

9.3.1 *Estimating* $V_\tau(\tilde{s})$

As mentioned in Section 9.3, we need a way to estimate the value of $V_\tau(\tilde{s})$. When deciding to accept a request through the optimal policy as defined in Equation (9.11), it is unknown what state the system, $\tilde{s}$, is at time τ, for all τ. Since we do not know the state of the system, we do not know its corresponding value function. In this section, we propose a heuristic to estimate this value.

1. Solve the optimal policy of Equation (9.1), ignoring any cancellations in the optimization process.

2. Generate n different arrival processes different cancellation processes.

3. Simulate the acceptance of products in these processes using Equation (9.4), cancellations arrive according to the simulated cancellation process.

4. For every time unit t, $t = 1, \dots, T$ and simulation number k, $k = 1, \dots, n$ record the state of the system x and corresponding value function. Denote these as x_t^k and $V_t^k(x_t^k)$.

The heuristics are as follows. Note that in our notation, $V_t(x)$ represents the actual value function, as derived using Equation (9.1). On the other hand, V_t^k represents the recorded value of the value function of simulation k at time t. Next, let us define:

$$\mathbb{1}_{it} = \begin{cases} 1 & \text{if the system is in state } i \text{ at time } t,\ i = 0, 1, \dots, C;\ t = 1, \dots, T \\ 0 & \text{otherwise} \end{cases}$$

Next, define:

$$\chi_i(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{it}$$

$\chi_i(t)$ denotes the proportion of being in state i at time t. Let $\widehat{\chi(t)}$ represent the most likely state for a time value t that is, $\widehat{\chi(t)} = \max(\chi_1(t), \chi_2(t), \dots)$. Using this notation, we have defined the following heuristics:

$$
\begin{aligned}
V_t^{H_1}(x) &= \frac{1}{n}\sum_{k=1}^{n} V_t^k(x_t^k) && \forall t && (9.14)\\
V_t^{H_2}(x) &= med(V_t^k(x_t^k)) && \forall t && (9.15)\\
V_t^{H_3}(x) &= max(V_t^k(x_t^k)) - min(V_t^k(x_t^k)) && \forall t && (9.16)\\
V_t^{H_4}(x) &= V_t(\left\lfloor \frac{C}{2} \right\rfloor) && \forall t && (9.17)\\
V_t^{H_5}(x) &= \sum_{i=1}^{C} \chi_i(t) V_t(x_i) && \forall t && (9.18)\\
V_t^{H_6}(x) &= V_t(\bar{\chi}(t)) && \forall t && (9.19)\\
V_t^{H_7}(x) &= \frac{1}{6}\sum_{i=1}^{6} V_t^{H_i}(x) && \forall t && (9.20)\\
V_t^{H_8}(x) &= med(V_t^{H_1}(x), ..., V_t^{H_6}(x)) && \forall t && (9.21)\\
V_t^{H_9}(x) &= V_t(1) && \forall t && (9.22)
\end{aligned}
$$

$V_t^{H_1}$ calculates the average over all n simulations, for a given time unit. $V_t^{H_2}$ is done in similar fashion, but takes the median. The $V_t^{H_3}$ heuristic takes the difference of largest and smallest value. $V_t^{H_4}$ does not use any simulated values, but instead takes the value function evaluated at half capacity. In heuristic $V_t^{H_5}$, we first obtain the proportions of being in a given state at time t, and using these as weights to obtain an estimate. These same proportions are used in heuristic $V_t^{H_6}$, but rather than weighing, we only take the most common value (statistical mode) as an estimate for the state we find ourselves in. Finally, heuristics $V_t^{H_7}$ and $V_t^{H_8}$ take the average and median, respectively, over the first six heuristics.

Note that in these heuristics, we have defined two approaches: estimating the value function directly, or estimating the state first, then plugging in the result into a value function.

9.4 RESULTS

This section is organized as follows. First, we discuss the setup of our simulation in Section 9.4.1. Next, we discuss the performance of the different heuristics H in Section 9.3.1. Having identified what heuristic to use in our calculations, we present the results in 9.4.3. We will review the robustness of the model against the different scenarios in this same section. Next, three examples are given in Section 9.4.4, where we compare the DPC and $DPID$ methods.

9.4.1 *Simulation setup*

In this section, we will discuss the setup of our simulation. Revenues will be compared between five scenarios:

1. **base**: this is the baseline scenario. Cancellations are calculated using the approach described in [21]. Arrival processes are simulated using "perfect" demand estimates; that is, the same $Poisson(\lambda)$ process is used to forecast demand and to simulate.
2. **CxlEarlier**: in this scenario, we optimize according to the **base** scenario, but in the arrival processes cancellations occur earlier than expected.
3. **CxlLater**: this is the opposite scenario: in this case, we simulate cancellations that occur later than planned.
4. **FcOver**: when constructing the optimal policy, we purposely overforecast by 20%.
5. **FcUnder**: similarly, in this scenario we purposely underforecast by 20%.

The **base** case contains a like-for-like comparison between the $DPID$ and DPC method. Next, we have also developed four different scenarios that measure the sensitivity of the $DPID$ and DPC methods with respect to incorrect estimates of cancellation times and levels of demand. The **CxlEarlier** and **CxlLater** scenarios were constructed to identify robustness of $c_j(t, \tau)$ in the optimization. We proposed this as further research in [21]. The **FcOver** and **FcUnder** scenarios are used to study how the model performs when forecasts are incorrect. Combinations of these are not studied, so we can isolate the effects of misjudgements in either cancellation probabilities or forecasts.

The unconstrained demand factors, the ratio of unconstrained demand forecast and aircraft capacity, are given in Table 9.1.

Type	Base	CxlEarlier	CxlLater	FcOver	FcUnder
Scenario	1.09	1.09	1.09	1.31	0.87

Table 9.1: Demand factors for different scenarios

Table 9.1 shows that the base case has sufficient (mean) demand to fill the plane at 109%. These demand factors are kept for the **CxlEarlier** and **CxlLater** scenarios. For the **FcOver** and **FcUnder** scenarios, we have scaled the base demand factor by 1.2 and 0.8, respectively. This results in scenarios with demand factors of 1.31 and 0.87. For these scenarios, we assume cancellation occur according to the base case.

The (unconstrained) demand distribution is given in Table 9.2. Unconstraining is done through the framework outlined in [19]. The fares are also shown.

Class	DCP1	DCP2	DCP3	DCP4	DCP5	DCP6	DCP7	DCP8	DCP9	Total	Fare
1	0.24	0.10	0.05	0.05	0.05	0.10	0.05	0.19	0.19	0.19	1000
2	0.25	0.12	0.09	0.06	0.06	0.09	0.12	0.16	0.03	0.29	750
3	0.21	0.12	0.12	0.09	0.09	0.09	0.09	0.09	0.09	0.51	500

Table 9.2: Distribution of demand by time and fares for different classes. DCPs (data collection points) range from 1 (earliest) to 9 (just before departure).

From Table 9.2, it becomes evident that the majority of the demand is expected to come from class 3. Most of this demand comes in earlier DCPs (1, 2 and 3). Demand from class 2 is also expected earlier one, but closer to departure, in DCPs 7 and 8, 28% of its demand is forecast. Class 1 has a similar demand curve, but the proportion of demand is expected to arrive even closer to departure in DCPs 8 and 9. The last column displays the fares: these are 1000, 750 and 500. The data used for our simulations is based on real data, but we have aggregated demand up to fare family level. This not only helps us obtain better demand estimates, but it will also enable us to study the effects better. To stay consistent with literature, we use the terminology of "class".

To speed up simulations, we have chosen a capacity of 50 seats. $n = 500$ simulations were performed for each scenario and for different heuristics H, H_1 through H_9. Demand is assumed to follow a Poisson process. Time is discretized in such a way that the probability of having two booking requests in the same time unit, is chosen at $\epsilon = 0.001$. As the demand grows, we need more time units to satisfy this condition. The control mechanism, checking the (adjusted) fare against the bid price, is achieved by first finding the closest time unit, and looking up the bid price for this unit of time. Finding the closest time unit is achieved by first converting both the time units of both the reference (simulation demand) and target (bid price) vectors to absolute time, finding the closest match, and converting the index of that closest match back to the time unit of the target vector. In case of a duplicate match, the earlier unit of time is used.

9.4.2 *Robustness on H*

Before we look into the model's results, we will discuss the robustness of H. Recall that we use H to estimate future states, which depend on an a-priori unknown state of the system x_t. We refer to Equations (9.14) through (9.22) in Section 9.3.1 to gain an understanding on how these were constructed. After running 500 simulations, the estimates are as shown in Figure 9.1.

As time draws closer, the gradient estimate, as shown in Equation (9.3), declines. This makes sense, since the DPID formulation, from Equation (9.1), is non-increasing in both t and x. For a proof, we refer the reader to Talluri and Van Ryzin [65].

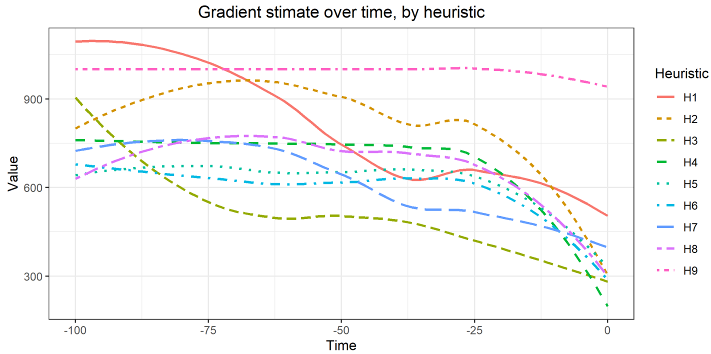


Figure 9.1: Comparison of different ways H_1 through H_9 to estimate V_t

Consider heuristic H_1. This line represents the gradient over time, averaged over all simulations. In absence of cancellations, this means on average the lowest price fare class, 3, start to become available only close to departure. The most aggressive heuristic, H_9, is equal to the highest fare, 1000, until close to departure when it starts to decrease. This again, is as expected: this heuristic measures last-seat availability. In effect, we impose the risk of having a final seat for sale for every booking we decide to accept or not. We will not go into detail how to interpret other heuristics as these are impossible to intuitively gain an understanding how these are constructed.

The most import part of this heuristic, is, of course the revenue it generates. We compare the revenues by fare class in Table 9.3. This is the result of 500 different arrival processes, which are being evaluated against every value function through simulation.

Class	H1	H2	H3	H4	H5	H6	H7	H8	H9
1	1.000	0.989	0.989	1.000	0.987	0.995	0.989	0.978	0.992
2	1.000	0.980	1.000	0.980	0.993	1.000	0.993	1.000	1.000
3	1.000	1.003	0.999	1.002	1.002	0.999	1.002	1.004	1.003

Table 9.3: Relative revenue performance of H by class, scaled to H_1

Table 9.3 shows the relative revenues scaled to the results obtained by H_1. We note very few differences between the different heuristics.

None of the heuristics increases class 1 sales. At best, this is matched by H_4, but this heuristic performs worst out of all heuristics at selling class 2. Heuristic H_8 is able to sell more customers to class 3, but does not sell class 1 well: 2.2% less than H_1. In Table 9.4, we show the total relative revenues.

H1	H2	H3	H4	H5	H6	H7	H8	H9
1.000	0.997	0.997	0.999	0.997	0.998	0.998	0.997	1.000

Table 9.4: Relative revenue performance of H, total, scaled to H_1

Table 9.4 analyzes the (total) mean revenue generated using heuristics H_1 through H_9. Note that H_1 performs best. Interestingly, the most aggressive heuristic, H_9 performs just as well. H_5 performs worst. Looking at Table 9.3, we observe that this is the result of selling less class 1 and 2. However, we observe that the average revenues performance are very close: using the worst heuristic we obtain revenues that are 0.3% lower than the best performing heuristic. We, therefore, conclude that our method is very robust against choice of heuristic. Considering that H_1 performs best, we have chosen to use this heuristic for the results we will show in the next section.

9.4.3 *Model results*

In this section, we review the performance of our formulation. We have divided this section in two parts: the first part, Section 9.4.3.1, covers the performance in terms of revenues and accepted passengers. The second part, Section 9.4.3.2, looks into the underlying processes and investigates customer behavior and compares this to the traditional DP formulation, *DPID*.

9.4.3.1 *Revenue*

Having identified our choice of H, we now present our results. We will review the accepted passengers by method and scenario, and then investigate the differences in the number of passengers accepted by class. Table 9.5 shows the revenues.

Scenario	Base	CxlEarlier	CxlLater	FcOver	FcUnder
DPID	29075	31450	29250	25000	30975
DPC	29875	30975	31950	29725	30725
Performance	1.03	0.98	1.09	1.19	0.99

Table 9.5: Revenue performance for different scenarios, comparing *DPID* with *DPC*.

Table 9.5 show the performance between *DPID* and *DPC* methods. For the base case, our algorithm outperforms the standard formulation by 3%. Looking at the performance when we misjudge cancellation probabilities, our model performs worse if cancellations happen earlier than we expect, and performs better if cancellations happen later than expected. We note that the performance is not linear: we lose 2% if they occur earlier, but gain 9% if they happen later. In When looking at the forecasting scenarios, we show a significant improvement over the *DPID* method when we overforecast, with a slight revenue loss when we underforecast.

When comparing the robustness of revenues between scenarios, we note a better performance for the **CxlEarlier** method, a 8.2% difference for *DPID* and 3.7% for *DPC*. For **CxlLater**, the *DPID* method is more robust: 0.6% compared to 6.8%. When overforecasting, *DPC* is more robust: 0.5% against 16% difference for *DPID*. Finally, when underforecasting, the robustness of revenues of *DPC* are again more favorable: 2.9% in comparison to 6.5% of *DPID*. In summary, we show minimal revenue losses in two out of five scenarios, and a much more robust revenues.

In Table 9.6 we look into the distribution of accepted passengers. Comparing our method with the traditional formulation, on average we accept slightly less class 1 passengers. This is consistent across the different scenarios, except the **CxlLater** scenario. Looking at class 2, we see a sharp decrease in number of accepted passengers. This seems to indicate that the value proposition, its fare and the risk of cancellation and taking up a unit of capacity, is not worth it.

To illustrate why this may happen, consider again Table 9.2. Here, we see a substantial amount of demand in DCP 5 through DCP 9 (46%). This roughly represents the last 25 days of the booking curve. Now consider Figure 9.2.

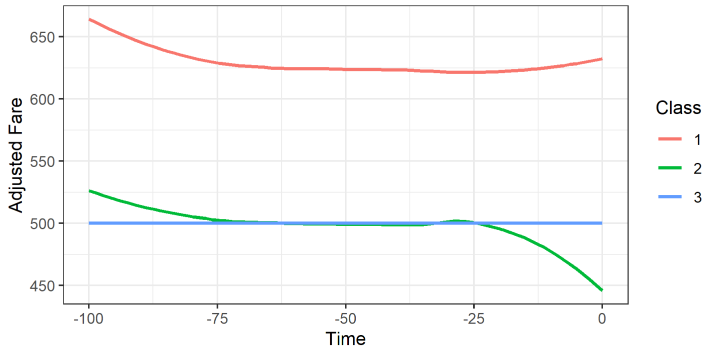


Figure 9.2: Adjusted fare by time of booking, by class

Figure 9.2 shows the adjusted fare by time of arrival. Here, we see that the adjusted fare for class 2 decreases below the value of class 3 (recall that class 3 is unable to cancel). Therefore, it is a better decision to accept class 3 at this stage. This behavior explains why the algorithm accepts more of class 2, and much more of class 3. We will investigate this behavior in more detail in the next section where we provide an example, Section 9.4.4.

Looking at the robustness between scenarios, we note that class 1 is the most robust and class 3 is the least robust, both in absolute and relative numbers. Class 2 show very consistent performance.

Lastly, we note very consistent total number of passengers in the *DPC* case: these range from 44.4 in the lowest case to 47.6 in the highest case: an absolute difference of 3.2 passengers. For the *DPID* method, this ranges from 32.7 to 38.9, a difference of 6.2 customers.

Class	1		2		3		Total	
Scenario	DPC	DPID	DPC	DPID	DPC	DPID	DPC	DPID
Base	10.30	11.20	10.10	21.10	24.00	4.10	44.40	36.40
FcOver	8.90	9.00	10.30	16.60	26.20	7.10	45.40	32.70
FcUnder	11.70	12.10	10.50	24.10	22.30	1.60	44.50	37.80
CxlEarlier	11.50	13.10	10.10	21.80	23.80	4.00	45.40	38.90
CxlLater	11.20	10.90	10.20	21.00	26.20	5.20	47.60	37.10

Table 9.6: Passenger count, by method and scenario

To understand the effects of a wrong cancellation prediction, consider Figure 9.3.

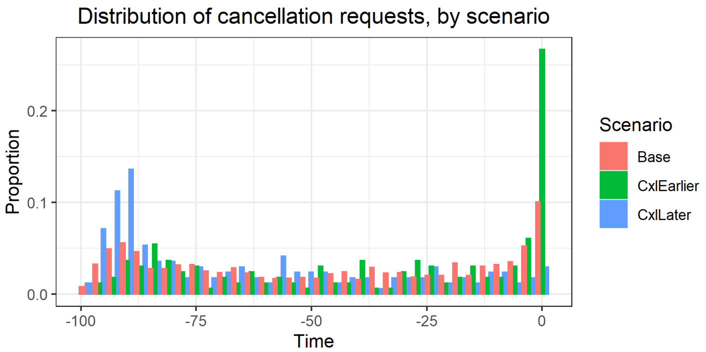


Figure 9.3: Example of wrong $c_j(t, \tau)$ curves

Figure 9.3 shows the distribution of cancellation rates for the **base**, **CxlEarlier** and **CxlLater** case for a booking made $t = 100$ days in advance. Note that we recognize the U shape we have seen in Chapter

7 and [21], with slightly more mass toward 0, indicating that this booking request is more likely to cancel late than early. For the **CxlLater** case, we put more mass on early cancellations in the arrival processes. Similarly, for the **CxlEarlier** case, we put more mass close to departure. This explains why. If we expect passengers to cancel late, in small values of t, we penalize bookings (since we have higher probabilities $c_j(t, \tau)$) harsher than we should have: we would have rejected someone. Similarly, if we expect passengers to cancel early, we do not penalize them as harsh. Moreover, we would have less booked passengers than we expected at early time units t, which will cause availability to open up more classes and we get a chance to accept more demand. We suspect this is why the revenue gains are not symmetric.

9.4.3.2 *Customer behavior*

In this section, we will review customer behavior. First, let us review the average number of cancellations. These are shown in Table 9.7.

Scenario	Base	CxlEarlier	CxlLater	FcOver	FcUnder
DPID	18.44	17.06	21.42	16.02	20.98
DPC	16.82	15.38	19.46	16.00	19.34
Gain	-1.62	-1.68	-1.96	-0.02	-1.64

Table 9.7: Average number of cancellations, by method and scenario

Table 9.7 shows that for every scenario, we report a lower number of cancellations. Keep in mind that the accepted number of passengers for *DPC* is higher than the *DPID* on average; refer to Table 9.6. Here, we see that the total number of passengers is about 8 larger for the *DPC* method. Despite having a larger number of passengers, we record a lower number of cancellations. This is true, except for the overforecasting, which shows a minimal gain. The number of cancellations is robust for the **CxlEarlier** and **FcOver** cases, compared to the **Base** case. However, in practice, an overall lower number of cancellations for the **Base** is important since this will make overbooking easier and less risky.

Looking into the underlying process into more detail, Table 9.8 shows the average number of requests that were cancelled, accepted and rejected. Note that the average number of accepted requests can exceed the capacity (in this case, 50), as long as there is capacity at the time of the request.

Type	Cancelled		Accepted		Rejected	
Scenario	DPC	DPID	DPC	DPID	DPC	DPID
Base	16.82	18.44	64.16	57.90	46.96	52.16
FcOver	16.00	16.02	62.74	51.92	26.30	35.80
FcUnder	19.34	20.98	66.24	62.00	66.76	70.22
CxlEarlier	15.38	17.06	60.36	54.36	50.02	55.80
CxlLater	19.46	21.42	68.22	62.42	49.44	53.34

Table 9.8: Average number of accepted, rejected and cancelled requests, by method and scenario

Table 9.8 shows the mean number of accepted and rejected requests. We have included the cancelled requests as well for sake of completion, this was earlier shown in Table 9.7. Note that the number of accepted requests is higher. This is the result of accepting more passengers (which we saw in Table 9.6) but, adjusted for those additional passengers, we seem to accept the right customers are the number of cancellations is lower. This leads to the observed cancellation rates, which we show in Table 9.9.

Scenario	DPC	DPID
Base	0.26	0.32
FcOver	0.26	0.31
FcUnder	0.29	0.34
CxlEarlier	0.25	0.31
CxlLater	0.29	0.34

Table 9.9: Mean observed cancellation rates, different scenarios

The observed cancellation rate in Table 9.9 is defined as the ratio of the number of cancellations and the number of accepted requests. Note that we observe less observed cancellation rates in all scenarios: this seems to indicate that the passengers we accept, are less likely to cancel. In reality, cancellations create uncertainty and angst in analysts that judge how much to overbook. For this reason this is another positive result in practice.

Table 9.10 shows the coefficient of variation of the observed cancellation rates, which is the ratio of the standard deviation and mean: $c_v = \frac{\sigma}{\mu}$.

Table 9.10 show the variability in the number of cancellations report. Looking at the cancellation requests, we observe a relatively higher variability for the **Base**, **FcUnder**, **FcOver** and **CxlLater** cases. Interestingly, **CxlEarlier** is the only scenario where the variability is

Type	Cancelled		Accepted		Rejected	
Scenario	DPC	DPID	DPC	DPID	DPC	DPID
Base	0.30	0.27	0.07	0.08	0.22	0.15
FcUnder	0.27	0.27	0.07	0.08	0.16	0.14
FcOver	0.23	0.19	0.05	0.08	0.35	0.20
CxlEarlier	0.23	0.26	0.05	0.07	0.21	0.16
CxlLater	0.25	0.23	0.07	0.08	0.22	0.17

Table 9.10: Coefficient of variation of cancellation rates, different scenarios. Smaller values represent better values.

lower, as compared to the *DPID* method. The number of accepted requests are more stable for all scenarios. The largest difference is in the number of rejected requests: for the **Base** case, the coefficient of variation is 46% higher. In other scenarios, such as **FcUnder**, this coefficient is variation is closer to the *DPID* case. We look into the number of rejected requests in Table 9.11.

Class	Base	CxlEarlier	CxlLater	FcOver	FcUnder
1	3.00	3.30	2.80	2.40	2.80
2	15.50	15.50	16.70	9.90	19.20
3	4.50	3.60	5.10	5.10	3.00

Table 9.11: Mean number of rejected requests in *DPC* that were accepted in *DPID*, different scenarios

In Table 9.11, we look at the average number of rejected requests, conditioning on the fact they were accepted in the *DPID* formulation. Particularly interesting are the rejected requests for class 1. For the base case, we reject, on average, three customers willing to buy class 1. However, looking back at Table 9.6, the final passenger count for *DPC* in the base case only had 0.9 less of class 1 booked. This seems to indicate that out of the 3 bookings that *DPID* accepted, on average, 2 of those bookings took up a valuable unit of capacity that was cancelled at some stage. We also confirm that the *DPC* model rejects a lot more of class 2 demand, that the *DPID* formulation did accept.

Figure 9.4 shows the distribution of the times arriving requests were rejected by scenario.

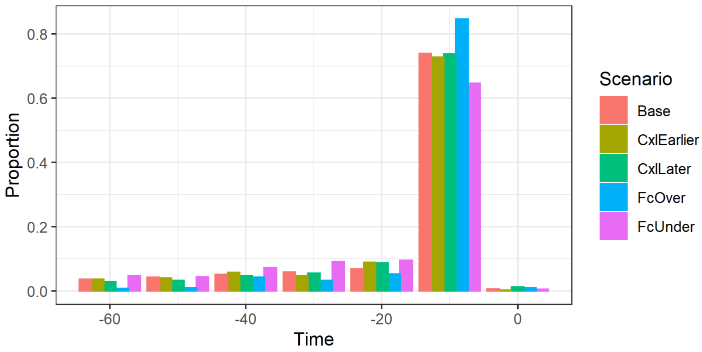


Figure 9.4: Distribution of time of rejection, by scenario

In Figure 9.4, we have bucketed time in multiples of 10 days, and calculated the proportion of rejections in that time bucket. Comparing the different scenarios, we first note very similar patterns across different scenarios. For all scenarios, the majority of rejections arrive between 10 and 0 days before departure. Intuitively, this makes sense: occupying a unit of capacity closer to departure is riskier to the airline that someone occupying a seat early in the booking curve.

9.4.4 *Example*

In this section, we will provide an examples of a simulation, comparing the performance of *DPID* and *DPC*. The revenue for *DPC* in this example is 31250, compared to the revenue of *DPID* of 2600. This represents a revenue improvement of 20%.

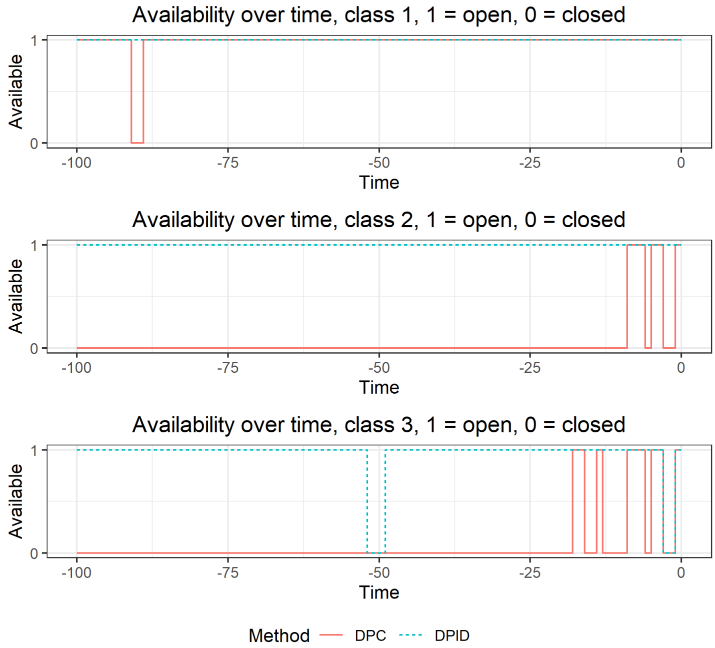


Figure 9.5: Example 1 - comparison of availability of classes over time, *DPC* vs *DPID*

Figure 9.5 shows the availability over time of this example for class 1, 2 and 3 in the top, middle and bottom graph. A value of 1 represents this class is available for sale, while 0 means this class was closed. We introduce a slight jitter to avoid overlapping lines in the figures. The solid line represents the availability for *DPC*, while the dotted line is the availability for *DPID*. A first interesting observation in the top graph is that class 1 is closed for 2 days, at $t = -91$ and $t = -90$. From the middle and bottom graphs, we see that these classes are also closed for the *DPC*. It is important to add that these classes were not closed, because capacity was exhausted. On the contrary, the formulation expects sufficient of demand later on, and this combined with the adjusted fare of class 1, that incorporates the estimated effect of taking up a unit of capacity, means it is optimal not to accept any class. While we only show one example in this section, we have seen this other simulations too, even for extended periods of time. Another example of the availability for class 1 in a different simulation is shown in Figure 9.7.

Looking at the availability of class 2, we note that this class is closed from the beginning of the booking curve for the *DPC* method, while it is available throughout of *DPID*. This is a phenomenon of the *DPID* that was reported earlier, in Chapter 8 or [22]. Now consider the time close to departure. Here, the algorithm opens class 2 for brief moments of time. Compare this availability with class 3 now: this class is open for a few days roughly two weeks before departure, while class 2 is closed. This is the result of the adjustment of fare for class 2, which we earlier highlighted with Figure 9.2. Figure 9.6 shows this in a different way by calculating the relative fare adjustment.

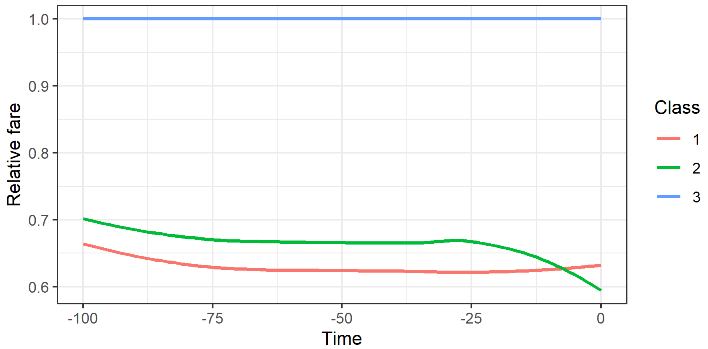


Figure 9.6: Relative fare adjustment by time of booking

From Figure 9.6, we observe that the relative fare adjustment drops closer to departure, and is strictly declining from roughty $t = -25$. The relative fare evaluated against the bid price drops from 68% to only 60% of its actual fare. This causes class 2 to be closed, while class 3 (that is not adjusted, since $c_j(t, \tau) = 0$ for all t) is open and is sold. This example shows non-nested availability in the last stages of the booking curve. This results in an increased number of rejected requests, as compared to the *DPID* model, which we show in Table 9.12.

	Class 1	Class 2	Class 3
Rejections	0	12	3

Table 9.12: Number of rejected requests in *DPC* that were accepted in *DPID*, by class

Table 9.12 shows the number of rejected requests that were accepted for the *DPID* model. Note that in this simulation, we reject arriving requests from class 2, a direct result of the aforementioned fare adjusted value.

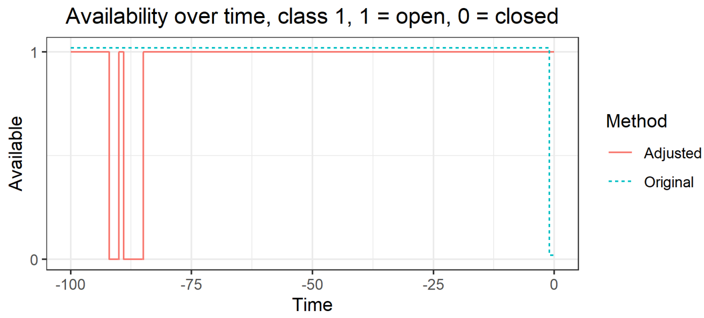


Figure 9.7: Example 2 - comparison of availability of class 1 over time, *DPC* vs *DPID*

9.5 DISCUSSION

In this section, we provide a discussion of the results and provide opportunities for further research.

In our earlier work, Chapter 7 and [21], we found that the time of cancellation is dependent on the time it was booked. Intuitively, this makes sense. However, in the literature, it is assumed a product cancels or not. It is seen as a Bernoulli process, which disregards time. In our earlier work, we provide a framework on how to estimate cancellation probabilities depending on t. However, while this modelling is important, it shouldn't be the objective itself: after all, the revenue that is generated using this model is what really matters.

In our formulation, we evaluate a product against the current opportunity cost, and the future opportunity cost of a unit of capacity weighted by the time-dependent. A-priori, we do not know the future state we are in. For this reason, we proposed different heuristics H. Estimating this term can be difficult, and there is no intuitive way to determine what heuristic of estimation is optimal. We have to rely on simulations to determine what method works best. These heuristics are based on the state of the system, at time t and simulation number i, $x_i(t)$. When we aggregate value function trajectories, such as H_1, it is impossible to calculate the gradient at that point. After all, the value function is defined at $x_i = 0, 1, ..., 50$, that is, the value function only exists at (whole) points of x_i. To approximate the gradient, we compared the value function estimate to the list of the true value function, find the associated x_i, and use this as a gradient. Suppose we find an estimate at some time t, with heuristic H_1. Let this estimate be $h(t)$. To approximate the gradient, we find:

$$\arg\min_{i} \left(abs(h(t) - V_t(x_i))\right) \tag{9.23}$$

Let i_t^* be the i that minimizes Equation (9.23). Then, we approximate by:

$$\frac{\partial H(t)}{\partial x} = \Delta V_t(x_{i^*}) \tag{9.24}$$

Note that this is an approximation that may impact the values of H. However, as we have shown in Section 9.4.2, the choice of this parameter is very robust against revenues. We therefore conclude that this approximation in finding the gradient is reasonable.

We have expressed f_j, adjusted for cancellation risk by f_j^*. We prefer to keep the adjustment term on the left-hand side of the equation, such as was done in Equation (9.12). This is preferred, since this enables airlines with the *DPID* formulation to easily implement our model, by using a fare adjustment. This feature, adjusting the fare that is used in the optimization, is present in all RM systems.

Looking at the results, we observe that our *DPC* model outperforms the *DPID* for most scenarios. We also stress that while the revenues increase, load factors increase too. This is important since in practice, lots of "professionals" still use load factors as a benchmark to see how a flight is performing. Even these days, analysts and managers alike are hard to convince that lower load factors may result in higher revenues. Fortunately, we accomplish both increases in revenues and load factors.

Looking at the different scenarios, misjudging time of cancellations, and over- or underforecasting, we found the *DPC* method to have relatively robust revenues and booking class distributions. This will, in the long run, ensure more stable forecasts. And, in turn, ensure more reliable value function estimates and corresponding bid prices.

It is not immediately clear how this method works in network RM. The currently most common method, which assumes that the network dynamic program is a sum of flight-level dynamic program, shown in this chapter, can easily be adapted to incorporate our work. Another way for further research is the robustness against multiple scenarios, different demand curves, different fare structures and more fare classes. Furthermore, it will be interesting to study structural properties of our dynamic programming formulation. Lastly, in Table 9.10, we showed the coefficient of variation. One finding we brought up in Section 9.4.3.2 is the relatively high coefficient of variation of rejected requests. It was not immediately clear why the number of rejected requests has a higher variance for the *DPC* model, and this is something that may be studied.

9.6 CONCLUSION

In summary, we have the following observations and conclusions:

1. Traditionally, optimisation methods that do assume cancellations, use probabilities that are class-independent (Section 2.6.1) or do not consider the time of booking. We have found this to be important in Chapter 7.
2. We introduce a novel dynamic programming method that uses the cancellation probabilities calculated in Chapter 7. This method is dependent on a heuristic that estimates the state of the system at time of cancellation.
3. Through simulation and nine different heuristics, we show how to estimate values of the system at future states, through a different dynamic programming formulation.
4. Using this heuristic, we keep a single-dimension state space in our novel dynamic programming method. This is important for fast run times, discussed in Chapter 4.
5. The optimal policy is found by adjusting the fare by the risk the airline faces of keeping one unit of capacity from sale.
6. By means of simulation, we show revenue increases between -2% and 19% and show how this method is robust against miscalculating cancellation probabilities and errors in forecasting.
7. We accept more passengers, but at the same time report a lower number of cancellations.
8. We show that it there are instances that it may be optimal to accept a non-refundable, lower-priced fare while rejecting a higher-priced fare which has the risk of cancelling. This means availability may no longer be nested.

ACRONYMS

AdvPur	Advance Purchase: the time in days between now and the flight departure date.
Alt	Alternative.
ASK	Available Seat Kilometer, total number of seats available for sale multiplied by distance.
Avl	Availability.
BP	Bid price: the opportunity cost.
Cls	Class.
Cxl	Cancellation.
DB	Denied Boarding.
DBD	Days Before Departure, the difference in days between now and date of departure.
DCP	Data Collection Point: a range of time over which data is collected.
DPC	Dynamic Programming with Cancellations.
DPDS	Dynamic Programming with Down Sell.
DPDS↑	Dynamic Programming with Down Sell and Waiting behavior.
DS	Double Exponential Smoothing.
EM	Expectation Maximisation.
EMSR	Expected Marginal Seat Revenue.
Fal	Fare of the host airline.
Fc	Forecast.
Fmkt	Fare of the cheapest competing, target airline.
GDS	Global Distribution System: making fares available to customers.
GP	Gaussian Processes.
KPI	Key Performance Indicator.
LF	Load Factor, number of occupied seats divided by total number of seats.
MNL	Multinomial Logit.
OB	Overbooking.
OD	Origin and Destination: the airport of origin and destination.
OTA	Online Travel Agent.

PD	Projection Detruncation.
PNR	Passenger Name Record: a booking.
PODS	Passenger Origin and Destination Simulator.
POS	Point of Sale: the country in which a ticket is sold.
RASK	Revenue per Available Seat Kilometer, revenue divided by available seat kilometers.
RM	Revenue Management.
Sch	Schedule.
Std	Standard Forecasting.
XGB	Extreme Gradient Boosting.

BIBLIOGRAPHY

[1] R.G. Cross. *Revenue Management*. Crown Business, New York, NY, 1997.

[2] LR Weatherford and PP Belobaba. Revenue impacts of fare input and demand forecast accuracy in airline yield management. *Journal of the Operational Research Society*, 53(8):811–821, 2002.

[3] Mark Brummer et al. Determination of potential load capacity based on observed load factor data. *A study for Northwest Airlines: St. Olaf College Undergraduate Practicum Group*, 1988.

[4] Peng Guo, Baichun Xiao, and Jun Li. Research on the EM algorithm in unconstraining estimation of non-normal distribution in revenue management. *Application of Statistics and Management*, 30(6):1077–1088, 2011.

[5] P. Belobaba and C. Hopperstad, editors. *Algorithms For Revenue Management In Unrestricted Fare Markets*, Cambridge, MA, 7 2004. INFORMS Section on Revenue Management.

[6] Dan Zhang and Larry Weatherford. Dynamic pricing for network revenue management: A new approach and application in the hotel industry. *INFORMS Journal on Computing*, 29(1):18–35, 2017.

[7] Guillermo Gallego and Ruxian Wang. Multiproduct price optimization and competition under the nested logit model with product-differentiated price sensitivities. *Operations Research*, 62(2):450–461, 2014.

[8] Una McMahon-Beattie. Trust, fairness and justice in revenue management: Creating value for the consumer. *Journal of Revenue and Pricing Management*, 10(1):44–46, 2011.

[9] Ian Seymour Yeoman and Una McMahon-Beattie. The turning points of revenue management: a brief history of future evolution. *Journal of Tourism Futures*, 2017.

[10] Daniel Adelman. Dynamic bid prices in revenue management. *Operations Research*, 55(4):647–661, 2007.

[11] Dan Zhang. An improved dynamic programming decomposition approach for network revenue management. *Manufacturing & Service Operations Management*, 13(1):35–52, 2011.

[12] Jiannan Ke, Dan Zhang, and Huan Zheng. An approximate dynamic programming approach to dynamic pricing for network revenue management. *Production and Operations Management*, 28(11):2719–2737, 2019.

[13] P. Belobaba. *Air Travel Demand and Airline Seat Inventory Management*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 1987.

[14] Patrick Gillespie. Judge throws out united airlines lawsuit against 22-year-old. https://money.cnn.com/2015/05/01/investing/united-airlines-lawsuit-skiplagged/index.html, May 1, 2015. *CNN*. Accessed March 13, 2019.

[15] Tara John and Nadine Schmidt. Lufthansa sues passenger who skipped his flight. Lufthansasuespassengerwhoskippedhisflight, February 13, 2019. *CNN*. Accessed March 13, 2020.

[16] Berlin District Court. District court summary 6 c15/18 (court document). https://www.franz.de/fileadmin/user_upload/urteil_ag_mitte_anonymisiert.pdf. Accessed February 1, 2019.

[17] Daniel Hopman, Ger Koole, and Rob van der Mei. Exploitative customer behavior in airlines: an overview of strategies for the smart customer. Working paper. Manuscript in preparation.

[18] Daniel Hopman, Ger Koole, and Rob van der Mei. Practical limitations of revenue management: theory versus practice. Working paper. Manuscript in preparation.

[19] Ilan Price, Jaroslav Fowkes, and Daniel Hopman. Gaussian processes for unconstraining demand. *European Journal of Operational Research*, 275(2):621–634, 2019.

[20] Daniel Hopman, Ger Koole, and Rob van der Mei. A machine learning approach to itinerary-level booking prediction in ten competitive airline markets. Working paper. Manuscript in preparation.

[21] Daniel Hopman, Ger Koole, and Rob van der Mei. A framework for modelling time-dependent cancellations in airline revenue management. Working paper. Manuscript in preparation.

[22] Daniel Hopman, Ger Koole, and Rob van der Mei. Single-leg revenue management with downsell and delayed decision making. *Journal of Revenue and Pricing Management*, 16(6):594–606, 2017.

[23] Daniel Hopman, Ger Koole, and Rob van der Mei. Leg-level dynamic programming with booking-time dependent cancellation probabilities. Working paper. Manuscript in preparation.

[24] Martin J Beckmann. Decision and team problems in airline reservations. *Econometrica: Journal of the Econometric Society*, pages 134–145, 1958.

[25] Kenneth Littlewood. Forecasting and control of passenger bookings. *Airline Group International Federation of Operational Research Societies Proceedings, 1972*, 12:95–117, 1972.

[26] Stefan Poelt. Practical pricing and the airline industry. In *Revenue Management*, pages 192–207. Springer, 2011.

[27] Arch G Woodside and Roberta MacDonald. General system framework of customer choice processes of tourism services. *Spoilt for Choice*, 30, 1994.

[28] Sara Dolnicar, Klaus Grabler, Bettina Grün, and Anna Kulnig. Key drivers of airline loyalty. *Tourism Management*, 32(5):1020–1026, 2011.

[29] Grahame R Dowling and Mark Uncles. Do customer loyalty programs really work? *Sloan Management Review*, 38:71–82, 1997.

[30] Maxwell Sandada and Bright Matibiri. An investigation into the impact of service quality, frequent flier programs and safety perception on satisfaction and customer loyalty in the airline industry in southern africa. *South East European Journal of Economics and Business*, 11(1):41–53, 2016.

[31] Mateja Kos Koklic, Monika Kukar-Kinney, and Spela Vegelj. An investigation of customer satisfaction with low-cost and full-service airline companies. *Journal of Business Research*, 80:188–196, 2017.

[32] Nikos Tsikriktsis and Janelle Heineke. The impact of process variation on customer dissatisfaction: Evidence from the us domestic airline industry. *Decision Sciences*, 35(1):129–141, 2004.

[33] Jin Woo Park, Rodger Robertson, Cheng-Lung Wu, et al. Investigating the effects of airline service quality on airline image and passengers' future behavioural intentions: findings from australian international air passengers. *Journal of Tourism Studies*, 16(1):2, 2005.

[34] Fredrik Carlsson and Åsa Löfgren. Airline choice, switching costs and frequent flyer programmes. *Applied Economics*, 38(13):1469–1475, 2006.

[35] Mohammed Naved Khan, Vippan Raj Dutt, and SC Bansal. Customer perceptions and expectations of service quality: A case study of domestic airline industry in india. In *Proceedings of the 2nd IIMA Conference on Research in Marketing*, pages 121–129, 2007.

[36] Ozlem Atalik and Melike Arslan. A study to determine the effects of customer value on customer loyalty in airline companies operating: case of turkish air travelers. *International Journal of Business and Management*, 4(6):154–162, 2009.

[37] Woohyun Cho. *The Impact of Airline and Customer Characteristics on Airline and Airport Choice*. PhD thesis, University of Maryland, 2012.

[38] Baker David Mc A. Service quality and customer satisfaction in the airline industry: A comparison between legacy airlines and low-cost airlines. *American Journal of Tourism Research*, 2(1):67–77, 2013.

[39] Timothy L Keiningham, Forrest V Morgeson III, Lerzan Aksoy, and Luke Williams. Service failure severity, customer satisfaction, and market share: An examination of the airline industry. *Journal of Service Research*, 17(4):415–431, 2014.

[40] Mukarramah Modupe Adeola and Sulaimon Olanrewaju Adebiyi. Service quality, perceived value and customer satisfaction as determinant of airline choice in nigeria. *International Letters of Social and Humanistic Sciences*, 20:66–80, 2014.

[41] Usman Khan and Noreen Khan. Customer satisfaction in airline industry. *International Proceedings of Economics Development and Research*, 76:63, 2014.

[42] Rahim Hussain, Amjad Al Nasser, and Yomna K Hussain. Service quality and customer satisfaction of a uae-based airline: An empirical investigation. *Journal of Air Transport Management*, 42:167–175, 2015.

[43] Colin Law. The study of customer relationship management in thai airline industry: A case of thai travelers in thailand. *Journal of Airline and Airport Management*, 7(1):13–42, 2017.

[44] T Manivasugen and Rista Nova. Factors affecting customer preferences while choosing a low cost airline. *Global Journal of Commerce and Management Perspective*, 2(3):19–26, 2013.

[45] Thomas Bieger, Andreas Wittmer, and Christian Laesser. What is driving the continued growth in demand for air travel? customer value of air transport. *Journal of air transport management*, 13(1):31–36, 2007.

[46] Thorsten Teichert, Edlira Shehu, and Iwan von Wartburg. Customer segmentation revisited: The case of the airline industry. *Transportation Research Part A: Policy and Practice*, 42(1):227–242, 2008.

[47] Rohit Verma and Gerhard Plaschka. The art and science of customer-choice modeling: Reflections, advances, and managerial implications. *Cornell Hotel and Restaurant Administration Quarterly*, 44(5-6):156–165, 2003.

[48] Rohit Verma. Customer choice modeling in hospitality services: A review of past research and discussion of some new applications. *Cornell Hospitality Quarterly*, 51(4):470–478, 2010.

[49] William G Brunger. The impact of the internet on airline fares: The 'internet price effect'. *Journal of Revenue and Pricing Management*, 9(1-2):66–93, 2010.

[50] Jose D Pérezgonzález and Andrew Gilbey. Predicting skytrax's official world airline star ratings from customer reviews. *Aviation Education and Research Symposium*, 2011.

[51] Tina Boetsch, Thomas Bieger, and Andreas Wittmer. A customer-value framework for analyzing airline services. *Transportation Journal*, 50(3):251–270, 2011.

[52] Christine SM Currie. How far will the airline model stretch? *Journal of Revenue and Pricing Management*, 14(2):120–122, 2015.

[53] Gregory M Coldren, Frank S Koppelman, Krishnan Kasturirangan, and Amit Mukherjee. Modeling aggregate air-travel itinerary shares: logit model development at a major us airline. *Journal of Air Transport Management*, 9(6):361–369, 2003.

[54] Gustavo Vulcano, Garrett Van Ryzin, and Wassim Chaar. Om practice—choice-based revenue management: An empirical study of estimation and optimization. *Manufacturing & Service Operations Management*, 12(3):371–392, 2010.

[55] Shanna Trichês Lucchesi, Christine Tessele Nodari, Ana Margarita Larrañaga, and Luis Afonso dos Santos Senna. Airline customer loyalty through analysis of stated preference. *Journal of Transport Literature*, 9(2):25–29, 2015.

[56] Christina P Milioti, Matthew G Karlaftis, and Eleni Akkogiounoglou. Traveler perceptions and airline choice: A multivariate probit approach. *Journal of Air Transport Management*, 49:46–52, 2015.

[57] Yuan-Ho Chen, Ming-Lang Tseng, and Ru-Jen Lin. Evaluating the customer perceptions on in-flight service quality. *African Journal of Business Management*, 5(7):2854, 2011.

[58] Richard Ratliff and Guillermo Gallego. Estimating sales and profitability impacts of airline branded-fares product design and pricing decisions using customer choice models. *Journal of Revenue and Pricing Management*, 12(6):509–523, 2013.

[59] Alwin Haensel, Ger Koole, and Jeroen Erdman. Estimating unconstrained customer choice set demand: A case study on airline reservation data. *Journal of Choice Modelling*, 4(3):75–87, 2011.

[60] Alejandro Mottini and Rodrigo Acuna-Agost. Deep choice model using pointer networks for airline itinerary prediction. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1575–1583. ACM, 2017.

[61] Alix Lhéritier, Michael Bocamazo, Thierry Delahaye, and Rodrigo Acuna-Agost. Airline itinerary choice modeling using machine learning. *Journal of Choice Modelling*, 31:198–209, 2019.

[62] Priyanga Gunarathne, Huaxia Rui, and Abraham Seidmann. When social media delivers customer service: Differential customer treatment in the airline industry. *MIS Quarterly*, 42(2):489–520, 2018.

[63] Leslie M Bobb and Emre Veral. Open issues and future directions in revenue management. *Journal of Revenue and Pricing Management*, 7(3):291–301, 2008.

[64] Garrett J Van Ryzin and Jeffrey I McGill. Revenue management: Research overview and prospects. *Transportation Science*, 33(2):233–256, 1999.

[65] Kalyan T Talluri and Garrett J Van Ryzin. *The theory and practice of revenue management*, volume 68. Springer Science & Business Media, 2006.

[66] Peng Guo, Baichun Xiao, and Jun Li. Unconstraining methods in revenue management systems: Research overview and prospects. *Advances in Operations Research*, 2012, 2012.

[67] Larry R Weatherford. The history of unconstraining models in revenue management. *Journal of Revenue and Pricing Management*, 15(3-4):222–228, 2016.

[68] William L Cooper, Tito Homem-de Mello, and Anton J Kleywegt. Models of the spiral-down effect in revenue management. *Operations Research*, 54(5):968–987, 2006.

[69] Richard H Zeni. *Improved forecast accuracy in airline revenue management by unconstraining demand estimates from censored data*. PhD thesis, Rutgers, The State University of New Jersey, 2001.

[70] J Salch. Unconstraining passenger demand using the EM algorithm. In *Proceedings of the INFORMS Conference*, 1997.

[71] Craig Hopperstad. An alternative detruncation method. *Boeing Commercial Aircraft Company Internal Document*, 1995.

[72] Maya R Gupta and Yihua Chen. Theory and use of the EM algorithm. *Foundations and Trends in Signal Processing*, 4(3):223–296, 2011.

[73] Daniel Kew Skwarek. *Competitive impacts of yield management system components: forecasting and sell-up models*. PhD thesis, Massachusetts Institute of Technology, 1996.

[74] Richard Robert Wickham. *Evaluation of forecasting techniques for short-term demand of air transportation*. PhD thesis, Massachusetts Institute of Technology, 1995.

[75] Garrett J Van Ryzin and Jeffrey I McGill. Revenue management without forecasting or optimization: An Adaptive Algorithm for determining airline seat protection levels. *Management Science*, 46(6):760–775, 2000.

[76] Jerald F Lawless. *Statistical Models and Methods for Lifetime Data*. John Wiley & Sons, 2011.

[77] Patrick H Liu, Stuart Smith, Eric B Orkin, and George Carey. Estimating unconstrained hotel demand based on censored booking data. *Journal of Revenue and Pricing Management*, 1(2):121–138, 2002.

[78] Carrie Crystal Queenan, Mark Ferguson, Jon Higbie, and Rohit Kapoor. A Comparison of Unconstraining Methods to Improve Revenue Management Systems. *Production and Operations Management*, 16(6):729–746, 2007.

[79] Daniel Kew Skwarek. Revenue and traffic impacts of alternative detruncation methods. *Proceedings of AGIFORS Reservations and Yield Management Study Group. Zurich, Switzerland*, 1996.

[80] Craig Hopperstad. Passenger O/D Simulator–PODS V.6. In *AGIFORS Proceedings*, 1996.

[81] Larry R Weatherford. Unconstraining methods. *Proceedings of the AGIFORS Reservations and Yield Management Study Group*, 2000.

[82] Larry R Weatherford and Stefan Pölt. Better unconstraining of airline demand data in revenue management systems for improved forecast accuracy and greater revenues. *Journal of Revenue and Pricing Management*, 1(3):234–254, 2002.

[83] Martin J Beckmann and F Bobkoski. Airline demand: An analysis of some frequency distributions. *Naval Research Logistics Quarterly*, 5(1):43–51, 1958.

[84] R Martinez and M Sanchez. Automatic booking level control. In *AGIFORS Symposium Proc*, volume 10, 1970.

[85] Christiane Lemke, Silvia Riedel, and Bogdan Gabrys. Dynamic combination of forecasts generated by diversification procedures applied to forecasting of airline cancellations. In *Proceedings of the 2009 IEEE Symposium on Computational Intelligence for Financial Engineering*, pages 85–91. IEEE, 2009.

[86] Oren Petraru. *Airline passenger cancellations: modeling, forecasting and impacts on revenue management*. PhD thesis, Massachusetts Institute of Technology, 2016.

[87] Patrick H Liu. Hotel demand/cancellation analysis and estimation of unconstrained demand using statistical methods. *Revenue management and pricing: Case studies and applications*, pages 91–101, 2004.

[88] Christoph Hueglin and Francesco Vannotti. Data mining techniques to improve forecast accuracy in airline business. In *Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining*, pages 438–442. ACM, 2001.

[89] Jeroen de Korte and Sandjai Bhulai. Hospitality revenue management. Master's thesis, Vrije Universiteit Amsterdam, 2019.

[90] Dolores Romero Morales and Jingbo Wang. Passenger name record data mining based cancellation forecasting for revenue management. *Innovative Applications of OR*, 202(2):554–562, 2008.

[91] Nuno Antonio, Ana de Almeida, and Luis Nunes. Predicting hotel booking cancellations to decrease uncertainty and increase revenue. *Tourism & Management Studies*, 13(2):25–39, 2017.

[92] Nuno António, Ana de Almeida, and Luis MM Nunes. Using data science to predict hotel booking cancellations. In *Handbook of research on holistic optimization techniques in the hospitality, tourism, and travel industry*, pages 141–167. IGI Global, 2017.

[93] Nuno Antonio, Ana de Almeida, and Luis Nunes. Big data in hotel revenue management: Exploring cancellation drivers to gain insights into booking cancellation behavior. *Cornell Hospitality Quarterly*, 2019.

[94] Tsung-Hsien Tsai. A temporal case-based procedure for cancellation forecasting: a case study. *Current Politics and Economics of South, Southeastern, and Central Asia*, 20(2):159, 2011.

[95] Han-Chen Huang, Allen Y Chang, Chih-Chung Ho, et al. Using artificial neural networks to establish a customer-cancellation prediction model. *Przeglad Elektrotechniczny*, 89(1b):178–180, 2013.

[96] Adel Alaeddini, Kai Yang, Pamela Reeves, and Chandan K Reddy. A hybrid prediction model for no-shows and cancellations of outpatient appointments. *IIE Transactions on Healthcare Systems Engineering*, 5(1):14–32, 2015.

[97] Theodore Charles Botimer. *Airline pricing and fare product differentiation*. PhD thesis, Massachusetts Institute of Technology, 1994.

[98] Michael Zhang and Peter Bell. Price fencing in the practice of revenue management: An overview and taxonomy. *Journal of Revenue and Pricing Management*, 11(2):146–159, 2012.

[99] Zuo-Jun Max Shen and Xuanming Su. Customer behavior modeling in revenue management and auctions: A review and new research opportunities. *Production and Operations Management*, 16(6):713–728, 2007.

[100] Ben Vinod. The continuing evolution: Customer-centric revenue management. *Journal of Revenue and Pricing Management*, 7(1):27–39, 2008.

[101] Laurie Garrow, Mark Ferguson, Pinar Keskinocak, and Julie Swann. Expert opinions: Current pricing and revenue management practice across us industries. *Journal of Revenue and Pricing Management*, 5(3):237–247, 2006.

[102] Ian Yeoman and Una McMahon-Beattie. *Revenue Management: A Practical Pricing Perspective*. Springer, 2010.

[103] Gabriel Bitran and René Caldentey. An overview of pricing models for revenue management. *Manufacturing & Service Operations Management*, 5(3):203–229, 2003.

[104] E Andrew Boyd and Royce Kallesen. Practice papers: The science of revenue management when passengers purchase the lowest available fare. *Journal of Revenue and Pricing Management*, 3(2):171–177, 2004.

[105] B Vinod, CP Narayan, and RM Ratliff. Pricing decision support: Optimising fares in competitive markets. *Journal of Revenue and Pricing Management*, 8(4):295–312, 2009.

[106] Ben Vinod. The complexities and challenges of the airline fare management process and alignment with revenue management. *Journal of Revenue and Pricing Management*, 9(1-2):137–151, 2010.

[107] Dieter Westermann and John Lancaster. Improved pricing and integration with revenue management–the next step toward improved revenues. *Journal of Revenue and Pricing Management*, 10(3):199–210, 2011.

[108] Richard Ratliff and Ben Vinod. An applied process for airline strategic fare optimization. *Journal of Revenue and Pricing Management*, 15(5):320–333, 2016.

[109] Gary L Parker. *Merchandising-Ancillary Revenues, Fare Families, Branded Fares and "a la carte" Pricing*, volume 14. 4019822 Canada Incorporated, 2019.

[110] Sunmee Choi and Anna S Mattila. Impact of information on customer fairness perceptions of hotel revenue management. *Cornell Hotel and Restaurant Administration Quarterly*, 46(4):444–451, 2005.

[111] John Lancaster. The financial risk of airline revenue management. *Journal of Revenue and Pricing Management*, 2(2):158–165, 2003.

[112] Michael W Tretheway. Distortions of airline revenues: why the network airline business model is broken. *Journal of Air Transport Management*, 10(1):3–14, 2004.

[113] Stanislav Ivanov. *Hotel Revenue Management: From Theory to Practice*. Zangador, 2014.

[114] Breffni M Noone and Anna S Mattila. Hotel revenue management and the internet: The effect of price presentation strategies on customers' willingness to book. *International Journal of Hospitality Management*, 28(2):272–279, 2009.

[115] Robert G Cross, Jon A Higbie, and Zachary N Cross. Milestones in the application of analytical pricing and revenue management. *Journal of Revenue and Pricing Management*, 10(1):8–18, 2011.

[116] Tudor Bodea and Mark Ferguson. *Segmentation, Revenue Management and Pricing Analytics*. Routledge, 2014.

[117] Sunmee Choi and Anna S Mattila. Hotel revenue management and its impact on customers' perceptions of fairness. *Journal of Revenue and pricing Management*, 2(4):303–314, 2004.

[118] Jochen Wirtz and Sheryl E Kimes. The moderating role of familiarity in fairness perceptions of revenue management pricing. *Journal of Service Research*, 9(3):229–240, 2007.

[119] Sheryl E Kimes and Jochen Wirtz. Has revenue management become acceptable? findings from an international study on the perceived fairness of rate fences. *Journal of Service Research*, 6(2):125–135, 2003.

[120] Sheryl E Kimes and Jochen Wirtz. Perceived fairness of revenue management in the us golf industry. *Journal of Revenue and Pricing Management*, 1(4):332–344, 2003.

[121] Xuan Lorna Wang, Cindy Yoonjoung Heo, Zvi Schwartz, Patrick Legohérel, and Frédéric Specklin. Revenue management: progress, challenges, and research prospects. *Journal of Travel & Tourism Marketing*, 32(7):797–811, 2015.

[122] L.R. Weatherford. Survey results from airlines that use revenue management software systems. *Journal of Revenue and Pricing Management*, 8(4):323 – 329, 2009.

[123] Tak C Lee and Marvin Hersh. A model for dynamic airline seat inventory control with multiple seat bookings. *Transportation Science*, 27(3):252–265, 1993.

[124] Conrad J Lautenbacher and Shaler Stidham Jr. The underlying markov decision process in the single-leg airline yield-management problem. *Transportation Science*, 33(2):136–146, 1999.

[125] Janakiram Subramanian, Shaler Stidham Jr, and Conrad J Lautenbacher. Airline yield management with overbooking, cancellations, and no-shows. *Transportation Science*, 33(2):147–167, 1999.

[126] Abhuit Gosavii, Naveen Bandla, and Tapas K Das. A reinforcement learning approach to a single leg airline revenue management problem with multiple fare classes and overbooking. *IIE Transactions*, 34(9):729–742, 2002.

[127] Garrett Van Ryzin and Gustavo Vulcano. Simulation-based optimization of virtual nesting controls for network revenue management. *Operations Research*, 56(4):865–880, 2008.

[128] W. L. Cooper, T. Homem-de Mello, and A.J. Kleywegt. Models of the spiral-down effect in revenue management. *Operations Research*, 54(5):968–987, Sep 2006.

[129] R. Phillips. Efficient frontiers in revenue management. *Journal of Revenue and Pricing Management*, 11(4):371–385, 2012.

[130] T. Fiig, K. Isler, C. Hopperstad, and S.K. Olsen. Forecasting and optimization of fare families. *Journal of Revenue and Pricing Management*, 11(3):322–342, 2012.

[131] T. Fiig, K. Isler, C. Hopperstad, and P. Belobaba. Optimization of mixed fare structures: Theory and applications. *Journal of Revenue and Pricing Management*, 9(1-2):152–170, 01 2010.

[132] JBG Frenk, Behrooz Pourghannad, and Semih O Sezer. A static model in single leg flight airline revenue management. *Transportation Science*, 51(1):214–232, 2016.

[133] Jochen Gönsch. A survey on risk-averse and robust revenue management. *European Journal of Operational Research*, 263(2):337–348, 2017.

[134] Dimitris Bertsimas and Ioana Popescu. Revenue management in a dynamic network environment. *Transportation Science*, 37(3):257–277, 2003.

[135] Victor Pimentel, Aishajiang Aizezikali, and Tim Baker. An evaluation of the bid price and nested network revenue management allocation methods. *Computers & Industrial Engineering*, 115:100–108, 2018.

[136] Syed Arbab Mohd Shihab, Caleb Logemann, Deepak-George Thomas, and Peng Wei. Autonomous airline revenue management: A deep reinforcement learning approach to seat inventory control and overbooking. *ICML 2019 Workshop*, 2019.

[137] Jiangang Dai, Anton J Kleywegt, and Yongbo Xiao. Network revenue management with cancellations and no-shows. *Production and Operations Management*, 28(2):292–318, 2019.

[138] K.T. Talluri and G.J. Van Ryzin. Revenue management under a general discrete choice model of consumer behavior. *Management Science*, 50, 2004.

[139] Q. Liu and G. van Ryzin. On the choice-based linear programming model for network revenue management. *Manufacturing & Service Operations Management*, 10(2):288–310, 2008.

[140] Dan Zhang and Daniel Adelman. An approximate dynamic programming approach to network revenue management with customer choice. *Transportation Science*, 43(3):381–394, 2009.

[141] S. Kunnumkal and H. Topaloglu. A new dynamic programming decomposition method for the network revenue management problem with customer choice behavior. *Production and Operations Management*, 19(5):575–590, 2010.

[142] Dan Zhang and Daniel Adelman. An approximate dynamic programming approach to network revenue management with customer choice. *Transportation Science*, 43(3):381–394, 2009.

[143] Juan José Miranda Bront, Isabel Méndez-Díaz, and Gustavo Vulcano. A column generation algorithm for choice-based network revenue management. *Operations Research*, 57(3):769–784, 2009.

[144] DD Sierag, GM Koole, Robert D van der Mei, JI Van der Rest, and Bert Zwart. Revenue management under customer choice behaviour with cancellations and overbooking. *European Journal of Operational Research*, 246(1):170–185, 2015.

[145] Dirk Sierag and Rob van der Mei. Single-leg choice-based revenue management: a robust optimisation approach. *Journal of Revenue and Pricing Management*, 15(6):454–467, 2016.

[146] Emmanuel Carrier and Larry Weatherford. Implementation of mnl choice models in the passenger origin-destination simulator (pods). *Journal of Revenue and Pricing Management*, 14(6):400–407, 2015.

[147] E. Carrier. Modeling airline passenger choice: Passenger preference for schedule in the passenger origin-destination simulator. Master's thesis, Massachusetts Institute of Technology, Cambridge, MA, 2003.

[148] L. Weatherford and R. Ratliff. Review of revenue management methods with dependent demands. *Journal of Revenue and Pricing Management*, 9, 2010.

[149] Tony SM Tse and Yiu-Tung Poon. Modeling no-shows, cancellations, overbooking, and walk-ins in restaurant revenue management. *Journal of Food Service Business Research*, 20(2):127–145, 2017.

[150] Anthony Sulistio, Kyong Hoon Kim, and Rajkumar Buyya. Managing cancellations and no-shows of reservations with overbooking to increase resource revenue. In *2008 Eighth IEEE International Symposium on Cluster Computing and the Grid (CCGRID)*, pages 267–276. IEEE, 2008.

[151] Red Ventures Company. Expertflyer: Expertflyer will change the way you fly. `https://www.expertflyer.com`, February 13, 2019. *CNN*. Accessed on multiple occasions between December 2016 and March 2020.

[152] Daniel P Kaplan and Mark R Dayton. Policies for the deregulated airline industry. *United States Congress*, 1988.

[153] Jyotsna Tyagi. Airline ticket booking. *International Journal of Advanced Technology in Engineering and Science*, 2, 2014.

[154] Lloyd Misquitta. Demystifying NDC. *SSRN*, 2019.

[155] Norbert Wiener. *Extrapolation, interpolation, and smoothing of stationary time series*, volume 7. MIT press Cambridge, MA, 1949.

[156] Anthony O'Hagan and John F C Kingman. Curve fitting and optimal design for prediction. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–42, 1978.

[157] Carl Edward Rasmussen and Christopher K I Williams. *Gaussian Processes for Machine Learning*. MIT press Cambridge, 2006.

[158] Montserrat Fuentes and Adrian E Raftery. Model evaluation and spatial interpolation by Bayesian combination of observations with outputs from numerical models. *Biometrics*, 61(1):36–45, 2005.

[159] Pei Gao, Antti Honkela, Magnus Rattray, and Neil D Lawrence. Gaussian process modelling of latent chemical species: applications to inferring transcription factor activities. *Bioinformatics*, 24(16):70—-75, 2008.

[160] Marc G Genton. Classes of kernels for machine learning: a statistics perspective. *Journal of Machine Learning Research*, 2(Dec):299–312, 2001.

[161] Roman Garnett, Michael A Osborne, and Stephen J Roberts. Sequential Bayesian prediction in the presence of changepoints. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 345–352. ACM, 2009.

[162] Yunus Saatçi, Ryan D Turner, and Carl Edward Rasmussen. Gaussian process change point models. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 927–934, 2010.

[163] Carl Edward Rasmussen and Hannes Nickisch. Gaussian processes for machine learning (GPML) toolbox. *Journal of Machine Learning Research*, 11(Nov):3011–3015, 2010.

[164] Andrew Gordon Wilson and Ryan Prescott Adams. Gaussian Process Kernels for Pattern Discovery and Extrapolation. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, pages 1067–1075, 2013.

[165] Chengjun Liu. Gabor-based kernel PCA with fractional power polynomial models for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(5):572–581, 2004.

[166] Rolf Rossius, Gérard Zenker, Andreas Ittner, and Werner Dilger. A short note about the application of polynomial kernels with fractional degree in Support Vector Learning. *Machine Learning: ECML-98*, pages 143–148, 1998.

[167] Dominique Guillot, Apoorva Khare, and Bala Rajaratnam. Complete characterization of Hadamard powers preserving Loewner

positivity, monotonicity, and convexity. *Journal of Mathematical Analysis and Applications*, 425(1):489–507, 2015.

[168] Carl H FitzGerald and Roger A Horn. On fractional Hadamard powers of positive definite matrices. *Journal of Mathematical Analysis and Applications*, 61(3):633–642, 1977.

[169] Ed Rigdon. Not positive definite matrices: causes and cures. *Georgia State University Website*, pages 1–6, 1997.

[170] John D Croston. Forecasting and stock control for intermittent demands. *Journal of the Operational Research Society*, 23(3):289–303, 1972.

[171] Aris A Syntetos and John E Boylan. On the bias of intermittent demand estimates. *International Journal of Production Economics*, 71(1-3):457–466, 2001.

[172] Aris A Syntetos and John E Boylan. The accuracy of intermittent demand estimates. *International Journal of forecasting*, 21(2):303–314, 2005.

[173] Fotios Petropoulos, Spyros Makridakis, Vassilios Assimakopoulos, and Konstantinos Nikolopoulos. 'horses for courses' in demand forecasting. *European Journal of Operational Research*, 237(1):152–163, 2014.

[174] David V Hinkley. Inference about the change-point in a sequence of random variables. *Biometrika*, pages 1–17, 1970.

[175] Michèle Basseville and Igor V Nikiforov. *Detection of Abrupt Changes: Theory and Application*. Prentice Hall Englewood Cliffs, 1993.

[176] Heung Wong, B Q Hu, W C Ip, and J Xia. Change-point analysis of hydrological time series using grey relational method. *Journal of Hydrology*, 324(1):323–338, 2006.

[177] Bruce Western and Meredith Kleykamp. A Bayesian change point model for historical time series analysis. *Political Analysis*, 12(4):354–374, 2004.

[178] Udayan Khurana, Deepak Turaga, Horst Samulowitz, and Srinivasan Parthasrathy. Cognito: Automated feature engineering for supervised learning. In *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, pages 1304–1307. IEEE, 2016.

[179] Finn Årup Nielsen. A new anew: Evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint arXiv:1103.2903*, 2011.

[180] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[181] David Foster. xgboostexplainer. Version 0.1.

[182] Amadeus. How artificial intelligence is transforming revenue management for airlines. https://amadeus.com/en/insights/blog/airline-ai-ml-revenue-management. Accessed December 30, 2019.

[183] Donald L. Povsner. If fares drop after you buy a ticket.. https://news.google.com/newspapers?id=ytIiAAAAIBAJ&sjid=KYQDAAAAIBAJ&pg=7111%2C5485695, Jul 14, 1985. *The Deseret News*. Accessed March 13, 2016.

[184] Department of Transporation. Guidance on the 24-hour reservation requirement. https://www.transportation.gov/sites/dot.gov/files/docs/Notice_24hour_hold_final20130530_0.pdf. Accessed March 13, 2016.