

**ESTIMATION AND OPTIMIZATION PROBLEMS IN
REVENUE MANAGEMENT WITH CUSTOMER CHOICE
BEHAVIOR**

A Thesis
Presented to
The Academic Faculty

by

Weijun Ding

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Industrial and Systems Engineering

Georgia Institute of Technology
May 2017

Copyright © 2017 by Weijun Ding

ESTIMATION AND OPTIMIZATION PROBLEMS IN REVENUE MANAGEMENT WITH CUSTOMER CHOICE BEHAVIOR

Approved by:

Professor Anton Kleywegt, Advisor
School of Industrial and Systems
Engineering
Georgia Institute of Technology

Professor Jim Dai
School of Operations Research and
Information Engineering
Cornell University

Professor David Goldberg
School of Industrial and Systems
Engineering
Georgia Institute of Technology

Professor Brani Vidakovic
School of Industrial and Systems
Engineering
Georgia Institute of Technology

Professor Jim Xu
College of Computing
Georgia Institute of Technology

Date Approved: 2016-12-06

ACKNOWLEDGEMENTS

First of all, I want to express my deep gratitude to my advisor Professor Anton Kleywegt for the advice and support I received from him. Professor Kleywegt dedicated a lot of time on my research and was always available for discussion when I had questions about anything during the past five years. His broad knowledge and insightful comments made the discussions so enjoyable and I could always learn a lot from these conversations.

I also want to thank Professor Jim Dai for bringing me to Georgia Tech and offering me very useful help and advice even after he left Georgia Tech.

I am also grateful to Professor Jim Xu for introducing me to the interesting research field in communication networks. I learned a lot about how to quickly explore an unknown field and how to think creatively from the collaboration with Professor Xu.

Special thanks go to Professor David Goldberg and Professor Brani Vidakovic for their valuable comments about my research.

I would also like to thank the revenue management and pricing teams at China Eastern Airlines, SAS, and Sabre. The collaboration with China Eastern Airlines and the internships at SAS and Sabre greatly shaped my research topics and it is exciting to see our research results have the potential to be used in real world.

Finally, I want to thank my wife and parents for their love and care during the course of my Ph.D. study.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
LIST OF TABLES	vi
LIST OF FIGURES	vii
SUMMARY	viii
I INTRODUCTION AND LITERATURE REVIEW	1
1.1 Traditional Revenue Management	1
1.2 Discrete Choice Models	4
1.3 Revenue Management with Choice Models	6
1.3.1 Deterministic approximation methods	7
1.3.2 Approximate dynamic programming methods	9
II JOINT ESTIMATION OF CHOICE MODEL PARAMETERS AND ARRIVAL RATE	10
2.1 Introduction	10
2.1.1 The No-Purchase Dilemma	11
2.1.2 The Nonhomogeneous Arrival Dilemma	12
2.2 Literature Review	13
2.3 Estimation Problem and Identifiability Conditions	16
2.3.1 Homogeneous arrival process	17
2.3.2 Piecewise constant arrival rate	24
2.3.3 Any integrable arrival rate	28
2.4 Estimation Method	33
2.4.1 Estimation of V	34
2.4.2 Construction of cumulative arrival rate function	35
2.4.3 Convergence proof	36
2.5 Numerical Examples	45
2.5.1 Effects of variations of assortments	45

2.5.2	Effects of Purchase Rate	47
2.5.3	Estimation of nonhomogeneous arrival rate	48
2.6	Conclusion	51
III	REVENUE MANAGEMENT UNDER CHOICE MODEL WITH BUY-DOWN EFFECTS	53
3.1	Introduction	53
3.1.1	Contributions	57
3.2	Related literature	58
3.3	Dynamic Assortment Optimization with General Buy-down Effects .	61
3.3.1	MNL model with buy-down effects	61
3.3.2	The decision problem and its MDP formulation	62
3.3.3	CDLP formulation	63
3.3.4	SBLP formulation	65
3.4	Dynamic Assortment Optimization with 100% Buy-down Effects . .	71
3.5	Assortment Optimization with Buy-down Effects and Random No- purchase	75
3.5.1	Notation and the assortment optimization problem	76
3.5.2	Assortment optimization with 100% buy-down effects and ran- dom no-purchase alternative	78
3.5.3	Assortment optimization with general buy-down effects and random no-purchase	83
3.6	Conclusion	85
APPENDIX A	— PROOFS FOR CHAPTER 3	86
REFERENCES	97

LIST OF TABLES

1	Quality of estimation for nonhomogeneous arrival rates	52
---	--	----

LIST OF FIGURES

1	Quality of estimation VS. variation of assortments	46
2	Quality of estimation VS. purchase rate	47
3	Estimation of cumulative arrival rate.	50
4	Sales distribution when different fare classes are the cheapest available	55
5	Conversion of a feasible SBLP solution to a feasible CDLP solution.	69

SUMMARY

Revenue management is about the application of various analytical techniques to optimize the price and availability of a company's products so the revenue performance is optimized. To successfully achieve the goal of revenue maximization, the revenue managers must have an accurate understanding of the customer behavior. One clear trend in revenue management is to model customer behavior using discrete choice models. This thesis deals with some of the challenges that arise with the incorporation of the discrete choice models in revenue management.

The first part of the thesis studies the parameter estimation problem in revenue management with discrete choice models. Revenue management models that include customer choice behavior have among others two types of parameters: (1) customer arrival rates and (2) choice parameters. In most applications, revenue managers have access to censored arrival data only, because the data do not include potential customers who decided not to purchase, that is, no-purchase data are missing. An important question is under what conditions all the arrival rate and choice parameters are identifiable with such censored data. When the arrival process is homogeneous, it has been known that if the censored data contain choices among only one assortment, then arrival rate and choice parameters are not identifiable. We consider a setting with multiple assortments, in which case arrival rate and choice parameters may or may not be identifiable. We derive the necessary and sufficient conditions for the arrival rate and the choice parameters to be identifiable with censored data. When the arrival process is a non-homogeneous Poisson process, the identification of arrival rate and the choice parameters is possible only when we have multiple realizations

of history. We derive the necessary and sufficient conditions for the identification of the arrival rate function and the choice parameters. Surprisingly, the identification conditions are only slightly more complicated than those when the arrival process is a homogeneous Poisson process. Also, the identification conditions do not depend on any knowledge of the arrival rate function. Based on this observation, we propose a semiparametric estimation procedure to jointly estimate the arrival rate function and the choice parameters. Both the estimate of the cumulative arrival rate function and the choice parameters are proved to converge to the true quantities almost surely when the data increase. Numerical examples also show that the algorithm can accurately estimate the arrival process and the choice parameters.

The second part of the thesis focuses on the revenue management problem with buy-down effects. The buy-down effects refer to the phenomenon that a product becomes more attractive if it is the cheapest available within certain subset of the assortment set, than if it is not the cheapest available within that subset. The multinomial logit (MNL) model can be modified to reflect the buy-down effects. We consider the dynamic assortment optimization problem under discrete choice model with buy-down effects. We propose a sales based linear programming (SBLP) formulation as a deterministic approximation to the original stochastic problem. Both the number of the decision variables and the number of constraints in the SBLP formulation are polynomial of the number of products. Thus, it is a much more efficient model than the popular choice based deterministic linear programming (CDLP) formulation. We give an efficient algorithm that converts an SBLP solution to a CDLP solution. We then consider the extreme case of the buy-down effects, where in each subset of the assortment, the customers only consider the cheapest available product. An SBLP formulation is developed and an algorithm that converts an SBLP solution to a CDLP solution is given. Lastly, we consider the extension where the no-purchase alternative in the choice set is random. We propose a polynomial algorithm to solve

the assortment optimization with 100% buy-down effects and random no-purchase alternative. When there is general buy-down effects, we prove that the optimal solution to the static assortment optimization under discrete choice model with buy-down effects and random no-purchase alternative is nested by revenue within subsets. The nesting property allows us to reduce the assortment optimization problem with general buy-down effects to the one with 100% buy-down effects, for which our efficient polynomial algorithm can be used.

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

1.1 Traditional Revenue Management

Revenue management is about the application of various analytical techniques to optimize the price and availability of a company's products so the revenue performance is optimized. The concept of revenue management was mainly originated from the airline industry, where it is also often called yield management. After the deregulation act from the Civil Aviation Board in 1978, airline companies obtained the freedom of pricing its own products. This resulted in fierce competition and a careful management of the price of the seats on an airplane became vital to the survival and prosperity of an airline company. By developing a revenue management system, American Airlines was able to stand out during the war with low cost airlines like People Express Airlines in 1980s. It was estimated in [48] that the revenue management techniques generated \$1.4 billion for American Airlines over a three year period in 1990s. Seeing the great value of the revenue management, other industries where the variable cost of a product is almost zero or the variable cost is relatively small compared with the fixed cost, like hospitalities, car rental, and many others, quickly adopted the revenue management technique (see [12]). For the ease of exposition, this thesis will be stated mainly in the context of airline industry. But the results hold in other industries as well.

One big difference between revenue management techniques and other profit maximization methods is that the former often tries to charge different prices to different customers through the control of price or availability, even though the products sold are exactly the same. For example, passengers next to each other on an airplane

often find their tickets prices very different. The major fact that enables a revenue management system to charge different prices for the same seats on an airplane is that customers have different willingnesses to pay for the same seat. Because the variable cost of a seat is almost zero, selling a seat even at very low price is better than leaving the seat empty. When there are too many empty seats, the airline should lower the price to attract customers. However, offering low price may also cannibalize the potential revenue from customers with high willingness to pay. To strike a balance between these two conflicting considerations, the traditional revenue management makes two major assumptions:

1. Each customer only wants to buy the ticket of one specific fare class (price) and there is an independent stream of customers for each fare class;
2. The customers that want the cheaper fare class arrive at the system before those want the more expensive fare class.

With these two assumptions, [35] considers a even simpler model where there are only two fare classes. The more expensive fare class is referred as class 1 and the cheaper one as class 2. Since the demand for class 2 arrives first, it is only needed to determine how many seats to reserve for class 1. The break-even point is to reserve as many seats for class 1 so that the expected marginal revenue generated from the reserved seats is equal to the price of class 2. Assume the demand for class 1 is a random variable with distribution function $F(\cdot)$. Let p_1 and p_2 denote the price of the two classes. The reservation level (also known as protection level) x for class 1 is determined as following:

$$p_1(1 - F(x)) = p_2 \quad \Rightarrow \quad x = F^{-1} \left(1 - \frac{p_2}{p_1} \right). \quad (1)$$

If the total capacity is denoted as C , $C - x$ is the booking limit of fare class 2.

The idea of balancing the expected revenue from the higher fare class with the immediate revenue from the lower fare class was extended in [3, 4, 5] to handle the

case where there are more than two fare classes. The idea is that when considering the booking limit of one fare class, all of the more expensive fare classes are grouped into a new mixed fare class. The problem is then reduced to the two-product case as in [35]. The method is called expected marginal seat revenue (EMSR) methods and different grouping methods lead to different EMSR methods.

The EMSR type methods successfully captured the trade-off between lost revenue of pricing the seats too high and the cannibalization effects of pricing the seats too low. They quickly became the major optimization method in the revenue management industry and are still popular in practice even today. However, as the industry where the revenue management techniques are applied evolve, both the practitioners and researchers found it more and more difficult to justify the two assumptions based on which the EMSR type and many other classical revenue management methods were developed: the independent demand assumption and the sequentially low-to-high demand arrival order.

The independence assumption implies that each customer knows which specific product she wants and she will not switch to other products even if the desired one is unavailable. This assumption is reasonable when revenue management ideas were first developed in airline industry at 1970/1980s. At that time, the number of fare classes from an airline is usually small. The relatively large price gap between products makes it insensible for the leisure customers to buy the more expensive ticket. On the other hand, the airlines introduced many different restrictions to the different fare classes so that the high-valued customers cannot easily switch to the cheap tickets.

However, after several dozens of years, the airline industry is now very different and the independence assumption is becoming unrealistic. First of all, many traditional purchase restrictions in the airline industry have been removed and airlines may also offer cheap tickets even close to departure time when they think it is profitable to do so. In this case, customers who originally are willing to buy more expensive tickets

may switch to the cheaper tickets instead. Besides, nowadays the airlines tend to have more fare classes than before. These fare classes differ not only in prices, but also in other attributes like mileage gain, cancellation fee, change fee and so on. With the development of Internet, more and more tickets are now booked online. Customers who book tickets online can easily compare the many tickets from many different airlines and then pick the one that fits their needs best. With the increasing number of fare classes, the increasing complexity of fare class attributes, and the new shopping behavior, a customer rarely knows which specific product she wants before actually seeing the available options. If a customer does not know which product she wants to book, assuming low-to-high demand arrival order is also meaningless.

On the other hand, the advancement of the Internet technology makes more data available. Nowadays, in the airline industry, not only the information pertaining to the customer booking process is recorded, e.g. the fare class that was booked and the time when the booking was made, also the availability of the fare classes are stored in the database by constantly taking snapshots of the system status throughout the booking horizon. The increase of the data amount and the improvement of the data quality make more sophisticated demand models possible.

The limitations of the historical demand model and the better data availability lead to the adoption of discrete choice model in revenue management.

1.2 Discrete Choice Models

Discrete choice model is a mathematical model that describes decision makers' choice behavior when they are presented with finitely many different alternatives. The set of all available alternatives is called the offer set and the set of alternatives that the decision maker actually considers is called the choice set. In general discrete choice models, the offer set and the choice set are not necessarily the same. More specifically, the choice set could be a strict subset of the offer set. In this study, we only consider

the case when the choice set is the same as the offer set.

Discrete choice models have many different variants. Among them, the simplest one is the multinomial logit (MNL) model. Let A denote the offer set. In the MNL choice model, each product j in A is associated with an attractiveness parameter V_j which is a positive number. the attractiveness parameter describes the relative popularity of a product to the customers. Given the offer set and the attractiveness parameters, according to the MNL choice model, the decision maker will choose product j with probability:

$$P_{j:A} = \frac{V_j}{\sum_{j' \in A} V_{j'}}.$$

The derivation of the MNL choice probability can be found in [6].

In practice, the attractivenesses of products of products can also be linked with the attributes of the products by assuming

$$V_j = \exp(\beta^T x_j),$$

where x_j is the attribute vector of product j and β are the attribute parameters.

The MNL choice model is simple and easy to estimate (see [6, 53]). It is also widely used in practice. However, the MNL choice model is often criticized due to its independence from irrelevant alternatives (IIA) property. The IIA property says that for a specific decision maker, the ratio of the choice probabilities of any two alternatives (given that both of them are in the offer set) is not affected by the attractiveness of any other alternatives. One classic example that shows the unrealisticness of the IIA property is the red bus / blue bus problem. Many other choice models are more general and can overcome the IIA limitation, e.g., the nested logit model, the mixed logit model, the latent class logit model, and so on. For more details of these classical choice models, please refer to [6, 53].

In recent years, many new choice models are also proposed. [23] propose a general attraction model (GAM) where each product is associated with two attractiveness

parameters, rather than one as in MNL. One of the attractiveness parameters is used when the product is included in the offer set and the other is used when the product is not in the offer set. This is to reflect the fact that customers may switch to other channels / suppliers when the product is not offered here. It is shown that both the MNL and the NL model are special cases of the GAM. Thus, the GAM is a more general model and can (partially) avoid the IIA property. [8] develop a Markov chain choice model where each product is modeled as a state in the Markov chain. The transition matrix and the initial probabilities to each state are the parameters of the model. The substitution of products is modeled as state transition and the final absorbing state is the customer's choice. It is shown that the GAM can be modeled exactly using the Markov chain choice model. Thus, the latter is even more general than the GAM.

[19] propose using a collection of preference lists to represent customers' preference. Each preference list is associated with a nonnegative weight to represent the percentage of the population that have this preference list. According to this model, each customer chooses the highest-ranked alternative from the available ones. Although the number of the potential preference lists is factorial in the number of products, [19] show that a relatively small number of preference lists is usually good enough for modeling purpose. [54] develop a column generation algorithm to estimate the preference list model.

1.3 Revenue Management with Choice Models

In the airline industry, when customers book tickets, they will face different tickets of fare classes from different flights of different airlines. With a little reflection, it can be seen that such a decision process is well fit for a discrete choice model. [1, 2] are among the first to use the MNL choice model to characterize and predict the customer booking behavior in the airline industry. But the optimization model in [2] is quite

elementary and heuristic. The seminal paper [50] propose a revenue management model with general customer choice behavior. The optimization problem determines which assortment to offer in each period and each state. The problem is modeled as a dynamic programming (DP) model which can incorporate any discrete choice model. [50] also discuss the data requirement and the parameter estimation problem for the revenue management problem with choice model. The framework proposed in [50] quickly became the standard model in revenue management with customer choice model.

1.3.1 Deterministic approximation methods

Since the DP model is often too big to solve and the resulting policy is complicated to implement in practice, some approximation methods are usually needed. [22] propose a deterministic approximation model where the stochastic quantities are replaced with their expected values and the capacity and the demand are treated as continuous. The resulting linear programming model is called choice based deterministic linear programming (CDLP). In CDLP, the number of decision variables is equal to the number of potential offer sets, which is exponential in the number of products. Thus, a column generation algorithm is often needed. The subproblem of the column generation algorithm is an assortment optimization problem. According to the equivalence between separation and optimization in [27], if one can show that the assortment optimization problem with certain choice model can be solved efficiently, the CDLP with the same choice model can also be solved efficiently. In recent years, quite a few efforts have been devoted to the assortment optimization problem.

Both [36] and [22] show that the optimal solution to the assortment optimization problem with the MNL model is nested by revenue and thus the problem can be solved efficiently. [14] show that the assortment optimization problem under the MNL model with totally unimodular constraint can also be solved efficiently through

linear programming models. [45] develop an efficient algorithm for the assortment optimization under the MNL model with capacity constraint. [44] discuss the assortment optimization problem under the MNL model with random choice parameters. It is shown that although the assortment optimization problem is easy to solve under certain conditions, the problem is generally NP-hard. To solve the same problem, [37] give a branch-and-cut algorithm and [10] propose a greedy heuristic algorithm. [46] consider a robust version of assortment optimization problem when the parameters are in a compact uncertainty set and show that an efficient algorithm can be developed. When the consideration sets are nested to each other in the MNL model, [20] prove that the corresponding assortment optimization problem is NP-complete. A fully polynomial time approximation algorithm is then proposed. [15] give the necessary and sufficient conditions under which the assortment optimization with nested logit model is polynomially solvable. When the conditions are satisfied, [33] give an efficient algorithm to solve the assortment optimization problem with d -level nested logit model. [17] developed fully polynomial time approximation schemes (FPTAS) for assortment optimization problem with the nested logit model and the mixed MNL model. [18] consider the assortment optimization problem under Markov choice model with capacity constraint. [57] consider the assortment optimization problem under general attraction model with capacity constraint.

When the choice model is the general attraction model, [23] develop a sales based linear programming (SBLP) model to approximate the dynamic assortment optimization problem. In the SBLP model, the decision variables are the sales quantities of the products, thus the number of decision variables is only linear in the number of products. This is a significant improvement over the CDLP model. [23] also show that the CDLP and SBLP solutions can be converted from one to the other within polynomial time, thus, the two formulations are equivalent. [21] develop an SBLP model under the Markov chain choice model. A polynomial time algorithm is developed to convert

the sales solution to the assortment solution of CDLP.

1.3.2 Approximate dynamic programming methods

Another interesting stream of literature uses approximation dynamic programming (ADP, see [7, 41] for introduction) to solve the DP model of the dynamic assortment optimization problem. [59] use an affine function to approximate the value function and show that the resulting program provides a tighter bound than the CDLP formulation. Based on the affine approximation, the DP is then formulated as a linear programming model with a large number of constraints. [28] consider using piecewise-linear function to approximate the value function. [29] prove bounds on how much the affine approximation and piecewise-linear approximation can tighten the CDLP formulation. [58] consider a nonlinear approximation to the value function and propose a simultaneous dynamic programming approach to solve the resulting nonlinear problem.

CHAPTER II

JOINT ESTIMATION OF CHOICE MODEL PARAMETERS AND ARRIVAL RATE

2.1 Introduction

Estimation of demand models is a core part of revenue management. Most modern demand models consist of two major components: a model of the population of potential customers, and a model of the choice behavior of the customers in the population. For example, the model of the population of potential customers often takes the form of an arrival process, such as a Poisson process with time-dependent rate function $\lambda(t)$. The model of the choice behavior of the customers represented by the population model often takes the form of a discrete choice model, such as a multinomial logit model, specifying the fraction $P_{j:S}$ of customers in the population who would choose alternative j out of choice set S , often including a no-purchase alternative.

In some traditional revenue management systems the distinction between the population model and the choice model is trivial. These demand models consist of a separate arrival process for each alternative, as though customers make their choices before observing the set of available alternatives. In these models the arrival processes specify both the population of potential customers as well as the customer choices. In this chapter we focus on demand models in which there is a choice component in addition to a population component, and we address two major obstacles encountered when estimating demand models with a population component and a choice component.

2.1.1 The No-Purchase Dilemma

The first obstacle is that the sales data that can be obtained in practice usually do not include data about potential customers who chose the no-purchase alternative. These no-purchase customers include customers who would have chosen an alternative other than the no-purchase alternative if a different set of alternatives were offered to the customer, and therefore it is important for revenue management purposes to include no-purchase customers in the demand model and to accurately estimate the population including such customers and the choice behavior including the no-purchase alternative. Depending on the demand model, no-purchase customers may include customers who choose alternatives offered by competitors or no-purchase customers may include only those customers who decide not to buy anything at this time. This distinction is important when modeling in practice, but for the purposes of this chapter the distinction is not important.

Next we illustrate the no-purchase dilemma. Consider the following two settings in which a product is offered at a particular price: In setting A , 100 customers consider the product, 50 choose to purchase the product, and the other 50 decide that it is too expensive and choose not to purchase the product. In setting B , 1000 customers consider the product, 50 choose to purchase the product, and the other 950 decide that it is too expensive and choose not to purchase the product. In both settings the seller observes only the 50 customers who purchase the product, and the seller has no data about the other customers. Thus, based on the available data, the seller cannot distinguish between setting A with a population of 100 and a purchase fraction of 0.5, and setting B with a population of 1000 and a purchase fraction of 0.05. This dilemma has been pointed out many times, including by [50] and [56]. On the other hand, it has also been reported in the literature that demand models with both a population component and a choice component, including a no-purchase choice, had been estimated successfully with data that do not include no-purchase

observations; see for example [40]. That raises the question of what the necessary and sufficient conditions are for demand models to be identifiable, when the demand models include both a population component and a choice component, including a no-purchase choice, and whether the conditions can be satisfied with data that do not include no-purchase observations.

2.1.2 The Nonhomogeneous Arrival Dilemma

Another difficulty in demand model estimation is how to estimate a nonhomogeneous customer arrival process. Existing estimation methods in the literature either assume that the arrival process is homogeneous or assume that the time horizon has been partitioned into intervals so that the arrival process within each interval is homogeneous. The dilemma is that if arrivals of no-purchase customers are not observed, then such partitioning and arrival rate estimation are problematic because the observed sales rate is affected not only by the unobserved total arrival rate, but also by the sets of alternatives (assortments) offered during the time horizon. For example, consider two intervals A and B . During interval A , 100 customers per unit time arrive, and a particular assortment is offered, resulting in 50 customers per unit time who choose to purchase a product, and the other 50 choose not to purchase a product. During interval B , 1000 customers per unit time arrive, and another assortment is offered, resulting in 50 customers per unit time who choose to purchase a product, and the other 950 choose not to purchase a product. Note that the seller cannot estimate the different arrival rates based on the observed data. This dilemma can be regarded as a more general version of the no-purchase dilemma discussed above, especially when the intervals are not predetermined or the arrival rate is not piecewise constant.

The difficulties mentioned above are serious obstacles inhibiting the use of modern demand models in revenue management practice. This chapter makes the following contributions to dealing with these challenges:

1. We study the identifiability of demand models consisting of an arrival model and a choice model, when only data about the assortments offered and the sales during the time horizon are available. Specifically, we give necessary and sufficient conditions for the identifiability of demand models for the cases in which the arrival process is homogeneous, piecewise constant, and general nonhomogeneous.
2. For the cases in which the arrival process is homogeneous or piecewise constant, if the demand model is identifiable, then the use of maximum likelihood estimation to estimate the demand model is relatively straightforward and has well-established desirable properties. However, in the case in which the arrival process is general nonhomogeneous, even if the demand model is identifiable, maximum likelihood estimation leads to degenerate estimates of infinite arrival rates at the time points of observed arrivals and an arrival rate of zero at all other times. We propose an estimation algorithm that jointly estimates a general nonhomogeneous arrival process and a choice model. We show that if the demand model is identifiable, then the estimates produced by the algorithm converge to the true quantities almost surely.

2.2 Literature Review

[50] is the seminal paper that first suggested incorporating a discrete choice model of customer behavior into the revenue management. Although the revenue management with discrete choice model has a lot of advantages over the traditional revenue management with independent demand model, [50] do point out that the estimation of the arrival rate and the choice model parameters using the sales data is difficult. [50] divide the time horizon into small intervals and the arrival in each small interval is modeled by a Bernoulli random variable, which is a discrete approximation to the Poisson arrival process. The likelihood of the sales can then be expressed in

terms of the Bernoulli parameter and the choice model parameters. An expectation-maximization (EM) algorithm is proposed to maximize the likelihood function. [38] show that a simpler likelihood function can be formulated by expressing the sales within a period where the offer set is constant as a filtered Poisson random variable. Thus, the time horizon does not need to be discretized. [38] also show that an EM algorithm can be developed to solve the new likelihood function. Another big concern about the estimation problem [50] raised is that there could be multiple pairs of parameters that lead to the same likelihood. The identifiability of the arrival rate and the choice model parameters has been noticed since the proposal of revenue management with discrete choice model.

To overcome the potential non-identifiability problem, [56] suggest using the market share of the company so one unique estimate can be determined from the likelihood function. [56] then develop an efficient EM algorithm to solve the resulting likelihood function. However, the EM algorithm may not be efficient when there are multiple historical assortments, which is a typical case in practical revenue management settings. More importantly, the claim (in Section 3.3) that there are always a continuum of maxima to the likelihood function turns out to be too conservative. For example, [55] implemented the EM algorithm to an empirical revenue management problem and show that the parameters can be estimated. As the discussion in this chapter will show, the maximum likelihood estimator admits a unique maximizer under certain conditions. [42] develop an heuristic algorithm that uses the demand mass balance equations to recover the market level demand. The algorithm also uses the airline's market share information to determine the relative attractiveness of the no-purchase option.

Given the sales data, one can write out the conditional likelihood function of the purchase choices conditioning on that the customers will select product from airline's offer sets. By maximizing the conditional likelihood function, the relative popularity

among the airline’s own products can be determined. The main difficulty of estimating the arrival rate (or arrival size) and the choice model parameters lies in the fact that the relative attractiveness of the no-purchase option is hard to determine since the number customers who purchased nothing is unobserved. On the other hand, once the relative attractiveness of the no-purchase option is determined, together with the attractivenesses of the airline’s own products, the arrival rate can then be easily determined. Several methods of this spirit have been proposed. [40] proposed a two-step algorithm to estimate the arrival rate and the choice model parameters. The first step estimated the choice model parameters by maximizing the conditional likelihood function. Based on the choice model parameters, the likelihood of the observed numbers of historical purchases can then be written as a function of the no-purchase attractiveness and the arrival rate. [40] show that the likelihood function can be further reduced to a scalar function of the no-purchase attractiveness. The resulting optimization problem is then easy to solve. [49] also proposed a two-step algorithm where the first step is the same as that in [40]. [49] noticed that the ratio between the numbers of expected purchases of two intervals with different offer sets does not depend on the arrival rate. The ratio is a function of the no-purchase attractiveness. [49] suggested finding the no-purchase attractiveness by minimizing the difference between the expected ratio of sales and the actual ratio of sales. As pointed out in [49], the effectiveness of the two-step method depends on sufficiently rich data, enough variation in the offer sets, as well as some prior knowledge of the arrival rate. Neither [40] nor [49] specify the exact conditions under which the two-step algorithms will work.

Some other papers exclusively focus on the estimation of the choice model parameters and the no-purchase attractiveness using only the sales data. [39] shows that if the choice model involved is some special generalized extreme value models (e.g. a nested logit model) so that the no-purchase attractiveness is not canceled out in the

conditional likelihood and there is enough variation in the offer sets, then the parameters of the no-purchase option or other completely censored option can be identified using conditional log-likelihood method or EM algorithm.

Most existing methods that estimate both the arrival rate and the choice model parameters assume that the arrival rate is homogeneous or assume that the horizon is partitioned so that the arrival process within each interval is homogeneous. As we argued in Section 2.1.2, such partitioning could be problematic. [32] explicitly models the nonhomogeneous arrival process while estimating the arrival rate and the customer behavior using only the sales data. In [32], a parameterized arrival intensity function is assumed and the estimation of the arrival rate boils down to the determination of the parameters. The methods still requires the practitioners to have a good knowledge of the market level demand, which is actually unobservable. On the other hand, when the data are not censored, [30] proposed a nonparametric method to estimate any nonhomogeneous Poisson process by using one or more realizations of the historical arrivals. The idea introduced in [30] will be extended in our estimation method so any nonhomogeneous market level arrival process can be estimated.

2.3 Estimation Problem and Identifiability Conditions

In this section, we will discuss the identifiability conditions of the joint estimation of the arrival rate and the choice model parameters using only the sales data and the availability data. In statistics, the identifiability is defined as following ([31]).

Definition 1. *If X is distributed according to probability law \mathbb{P}_θ where $\theta \in \Theta$ is the parameter of the distribution, then θ is said to be identifiable on the basis of X in Θ if for any $\theta_1, \theta_2 \in \Theta$ with $\theta_1 \neq \theta_2$, $\mathbb{P}_{\theta_1} \neq \mathbb{P}_{\theta_2}$ (Two probability laws being different means they differ on a set of nonzero measure).*

If a parameter is unidentifiable, there is no consistent estimator for the parameters and no precise inference of the parameters can be made.

Throughout this study, we assume the customer choice behavior follows the multinomial logit (MNL) model. The customer arrival process is a Poisson arrival process. We will discuss three different scenarios of the rate function of the Poisson process regarding its time homogeneity. The simplest case is when the arrival rate is constant as adopted in [40], [50], and [56]. This is also the case where the confusion about the identifiability arises. We then discuss the case when the arrival rate is piecewise constant and when the arrival rate is any integrable nonnegative function.

2.3.1 Homogeneous arrival process

Let J denote the set of company's products. The company may offer any $A \subset J$ as the offer set. Let L denote the length of the sales horizon. Throughout the horizon, we observe N different assortments. Assortment A_n is offered for L_n time units for $n = 1, \dots, N$. We have $\sum_{n=1}^N L_n = L$. We make the following assumptions:

1. Arrival customers make purchase decisions according to an MNL choice model. The attractiveness of product j is V_j and is a constant positive number.
2. The no-purchase option is always available and has constant attractiveness. The attractiveness of no-purchase is normalized to 1.
3. $L_n > 0$ for every n .
4. Every product has been offered in some assortment.
5. In the setting where only sales data are observed, an empty assortment does not provide any information about the arrival rate or the product attractiveness. So without loss of generality, we assume $A_n \neq \emptyset$ for every n .

We denote the collection of the assortments as \mathcal{S} and $|\mathcal{S}| = N$. During the time interval L_n when A_n is offered, we observe C_{nj} purchases of product j for $j \in A_n$. A_n , L_n , and C_{nj} are the available information for the estimation problem.

To simplify the notation, we use $V(A)$ to denote the sum of the attractivenesses of products in A . That is, $V(A) = \sum_{j \in A} V_j$. Given A_n 's and L_n 's, the likelihood of observing C_{nj} is as (2).

$$\ell(C|V, \lambda) = \prod_{n=1}^N \left[\exp \left(-\lambda L_n \frac{V(A_n)}{V(A_n) + 1} \right) \frac{\left(\lambda L_n \frac{V(A_n)}{V(A_n) + 1} \right)^{\sum_{j \in A_n} C_{nj}}}{\left(\sum_{j \in A_n} C_{nj} \right)!} \frac{\left(\sum_{j \in A_n} C_{nj} \right)!}{\prod_{j \in A_n} C_{nj}!} \prod_{j \in A_n} \left(\frac{V_j}{V(A_n)} \right)^{C_{nj}} \right] \quad (2)$$

It has been pointed out in [56] that when $N = 1$, (V, λ) is unidentifiable on the basis of C_{nj} in $\{(V, \lambda) : V > 0, \lambda > 0\}$. To see this, when $N = 1$, the likelihood function is reduced to

$$\ell_1(C|V, \lambda) = \exp \left(-\lambda L_1 \frac{V(A_1)}{V(A_1) + 1} \right) \frac{\left(\lambda L_1 \frac{V(A_1)}{V(A_1) + 1} \right)^{\sum_{j \in A_1} C_{1j}}}{\prod_{j \in A_1} C_{1j}!} \prod_{j \in A_1} \left(\frac{V_j}{V(A_1)} \right)^{C_{1j}}.$$

Then $\ell_1(C|V, \lambda) = \ell_1(C|V', \lambda')$ for every C where $V' = \alpha V$ and $\lambda' = \frac{\alpha V(A_1) + 1}{\alpha(V(A_1) + 1)} \lambda$ with any $\alpha > 0$.

Thus, to make (V, λ) identifiable, we must at least observe two different assortments. Unfortunately, this requirement alone does not guarantee the identifiability of the parameters, either, as shown in the following example.

Example 1. Let $N = 2$, $A_1 = \{1\}$, $A_2 = \{2\}$. Then

$$\ell_2(C|V, \lambda) = \exp \left(-\lambda L_1 \frac{V_1}{V_1 + 1} \right) \frac{\left(\lambda L_1 \frac{V_1}{V_1 + 1} \right)^{C_{1,1}}}{C_{1,1}!} \exp \left(-\lambda L_2 \frac{V_2}{V_2 + 1} \right) \frac{\left(\lambda L_2 \frac{V_2}{V_2 + 1} \right)^{C_{2,2}}}{C_{2,2}!}.$$

Then for any (V, λ) , (V', λ') defined with any $k > 1$ as following has $\ell_2(C|V', \lambda') = \ell_2(C|V, \lambda)$ for every C :

$$\begin{aligned} \frac{V'_1}{V'_1 + 1} &= \frac{1}{k} \frac{V_1}{V_1 + 1} \Rightarrow V'_1 = \frac{V_1}{k(V_1 + 1) - V_1} \\ \frac{V'_2}{V'_2 + 1} &= \frac{1}{k} \frac{V_2}{V_2 + 1} \Rightarrow V'_2 = \frac{V_2}{k(V_2 + 1) - V_2} \\ \lambda' &= k\lambda. \end{aligned}$$

Actually, following the argument in Example 1, it can be shown that even for larger N , if no two offered assortments have products in common, (V, λ) are unidentifiable.

We next use the Definition 1 to find the necessary and sufficient conditions of the identifiability. When discussing the identifiability, we will also work with the log-likelihood function as (3).

$$\ell(C|V, \lambda) = \sum_{n=1}^N \left[\left(-\lambda L_n \frac{V(A_n)}{V(A_n) + 1} \right) + \sum_{j \in A_n} C_{nj} \log \left(\lambda L_n \frac{1}{V(A_n) + 1} \right) \right. \\ \left. - \log \left(\prod_{j \in A_n} C_{nj}! \right) + \sum_{j \in A_n} C_{nj} \log(V_j) \right] \quad (3)$$

Before giving the necessary and sufficient conditions of the identifiability, we first introduce some notation and definition.

Definition 2. *Given a collection of assortments \mathcal{S} , assortments A_1 and A_2 in \mathcal{S} are said to be communicating on \mathcal{S} if $A_1 \cap A_2 \neq \emptyset$ or there exists $A_{(1)}, A_{(2)}, \dots, A_{(k)} \in \mathcal{S}$ for some $k \geq 1$ such that*

$$A_1 \cap A_{(1)} \neq \emptyset, \quad A_{(1)} \cap A_{(2)} \neq \emptyset, \quad \dots, \quad A_{(k-1)} \cap A_{(k)} \neq \emptyset, \quad A_{(k)} \cap A_2 \neq \emptyset.$$

Definition 3. *A collection of assortment \mathcal{S} are said to be communicating if every two assortments in \mathcal{S} are communicating on it.*

The following technical result will be used in the proof of the sufficiency of the identifiability conditions.

Suppose $M \in \mathbb{R}^{m \times n}$, and $\mathbf{1}$ denotes a vector of ones with appropriate dimension. Define a system of equations as following.

$$Mx = \mathbf{1}, \quad x > 0 \quad (\text{I})$$

Define two alternative systems of $\pi \in \mathbb{R}^m$ as following.

$$\mathbf{1}^T \pi > 0, \quad M^T \pi \leq 0 \quad (\text{A.1})$$

$$\mathbf{1}^T \pi = 0, \quad M^T \pi \leq 0, \quad M^T \pi \neq 0 \quad (\text{A.2})$$

Theorem 1. *System (I) has no solution if and only if System (A.1) or System (A.2) has a solution.*

Proof of Theorem 1. When System (A.1) has a solution π , we can pre-multiply both sides of System (I) with π^T and we have:

$$\pi^T Mx = \pi^T \mathbf{1}. \quad (4)$$

According to the second part of the constraints of (A.1) and the requirement that $x > 0$, $\pi^T Mx \leq 0$. While the first part of the constraints of (A.1) says $\pi^T \mathbf{1} > 0$. Thus, we reach a contradiction from System (I). This means (I) does not have a solution.

Similarly, when System (A.2) has a solution, we can also reach a contradiction for System (I).

We now show when System (I) does not have a solution, System (A.1) or (A.2) have a solution. We introduce a relaxed system (5) to (I) as following and our discussion will be based on whether the relaxed system has solution or not.

$$Mx = \mathbf{1}, \quad x \geq 0. \quad (5)$$

1. Neither (I) nor (5) has a solution. Since (5) does not have a solution, according to Farkas lemma, (A.1) has a solution.
2. (I) does not have a solution but (5) has a solution. This means the following linear programming problem (6) has optimal objective value equal to 0. If the optimal objective value is negative, $x^* + \mathbf{1}y^* > 0$ and $x^* + \mathbf{1}y^*$ is a solution to (I), which contradicts our starting assumption.

$$\min_{x,y} : -y, \quad \text{s.t.: } Mx + M\mathbf{1}y = \mathbf{1}, \quad x \geq 0, \quad y \geq 0. \quad (6)$$

According to the strong duality theory, the dual to (6) as below is feasible and has optimal objective value equal to 0.

$$\max_{\pi} : \mathbf{1}^T \pi, \quad \text{s.t.: } M^T \pi \leq 0, \quad \mathbf{1}^T M^T \pi \leq -1.$$

Because of the second constraint $\mathbf{1}^T M^T \pi \leq -1$, the optimal solution π^* to the above dual problem also satisfies $M^T \pi \neq 0$. Thus, π^* is a solution to (A.2).

To conclude, when (I) does not have a solution, (A.1) or (A.2) have a solution. This finishes the proof. \square

Given the collection of observed assortments, \mathcal{S} , we define an incidence matrix $M^{\mathcal{S}}$ of dimension $|\mathcal{S}| \times |J|$ as $M_{nj}^{\mathcal{S}} = 1$ if $j \in A_n$ and $M_{nj}^{\mathcal{S}} = 0$ otherwise. The identification conditions of the estimation problem can be summarized as following.

Theorem 2. *Let \mathcal{S} be the collection of assortments observed in history. (V, λ) is identifiable if and only if the system $\{M^{\mathcal{S}}x = \mathbf{1}, x > 0\}$ has no solution.*

Proof. We first prove that when $\{M^{\mathcal{S}}x = \mathbf{1}, x > 0\}$ has a solution \bar{x} , (V, λ) is unidentifiable. To prove this, we just need to find two different sets of parameters (V, λ) and (V', λ') such that $\ell\ell(C|V, \lambda) = \ell\ell(C|V', \lambda')$ for every C . To achieve this, we pick $V = \bar{x}$ and λ to be any positive number. Set $V' = kV$ and $\lambda' = \frac{1}{2} \frac{k+1}{k} \lambda$ with any $k > 0$ and $k \neq 1$, then $(V, \lambda) \neq (V', \lambda')$ and $\ell\ell(C|V, \lambda) = \ell\ell(C|V', \lambda')$ for every C .

Now we prove when $\{M^{\mathcal{S}}x = \mathbf{1}, x > 0\}$ does not have a solution, (V, λ) is identifiable.

Suppose we have (V, λ) and (V', λ') such that $\ell\ell(C|V, \lambda) = \ell\ell(C|V', \lambda')$:

$$\begin{aligned} & \sum_{n=1}^N \left[-\lambda L_n \frac{V(A_n)}{V(A_n) + 1} + \sum_{j \in A_n} C_{nj} (\log \lambda - \log (V(A_n) + 1) + \log V_j) \right] \\ &= \sum_{n=1}^N \left[-\lambda' L_n \frac{V'(A_n)}{V'(A_n) + 1} + \sum_{j \in A_n} C_{nj} (\log \lambda' - \log (V'(A_n) + 1) + \log V'_j) \right]. \end{aligned} \tag{7}$$

Here, $V'(A) := \sum_{j \in A} V'_j$.

For (7) to hold for all C_{nj} 's, we must have the coefficients before each C_{nj} on both

sides to be the same:

$$\log \lambda - \log (V(A_n) + 1) + \log V_j = \log \lambda' - \log (V'(A_n) + 1) + \log V'_j, \forall n, \forall j \in A_n; \quad (8)$$

$$\Rightarrow \lambda \frac{V(A_n)}{V(A_n) + 1} = \lambda' \frac{V'(A_n)}{V'(A_n) + 1}, \forall n. \quad (9)$$

If products j and l are in the same A_n for some n , then according to (8), we have

$$\log \lambda - \log (V(A_n) + 1) + \log V_j = \log \lambda' - \log (V'(A_n) + 1) + \log V'_j,$$

$$\log \lambda - \log (V(A_n) + 1) + \log V_l = \log \lambda' - \log (V'(A_n) + 1) + \log V'_l.$$

Subtracting the second equation from the first one, we have

$$\log(V_j) - \log(V_l) = \log(V'_j) - \log(V'_l) \Rightarrow \frac{V_j}{V'_j} = \frac{V_l}{V'_l}, \forall j, l \text{ such that } j, l \in A_n \text{ for some } A_n. \quad (10)$$

If $(A_{n_1} \cap A_{n_2}) \neq \emptyset$, then (10) holds for every $j, l \in A_{n_1} \cup A_{n_2}$. Following this logic, it is easy to check that if \mathcal{S}_k is a collection of assortments that is communicating, then there exists $\beta(\mathcal{S}_k) > 0$ such that

$$\frac{V_j}{V'_j} = \beta(\mathcal{S}_k), \quad \forall j \in \cup_{A_n \in \mathcal{S}_k} A_n. \quad (11)$$

Since $\{M^{\mathcal{S}}x = \mathbf{1}, x > 0\}$ does not have a solution, according to Theorem 1, there exists $\pi \in \mathbb{R}^T$ such that one of the alternative systems as (A.1) and (A.2) holds.

1. When π satisfies (A.1):

$$\mathbf{1}^T \pi > 0, \quad (M^{\mathcal{S}})^T \pi \leq 0.$$

We can partition the collection of assortments in \mathcal{S} with nonzero π into $\mathcal{S}_1, \dots, \mathcal{S}_K$ such that

- \mathcal{S}_k are communicating for each k ;

- $\mathcal{S}_k \cup \{A\}$ is not communicating for any $A \in \mathcal{S}_{k'}$ with $k \neq k'$.

Thus, \mathcal{S}_k 's are maximal communicating collection of assortments. We use π^k to denote the portion of π corresponding to \mathcal{S}_k . Then we have $(M^{\mathcal{S}_k})^T \pi^k \leq 0$ for all k and $\mathbf{1}^T \pi^k > 0$ for at least one k . Without loss of generality, we assume $\mathbf{1}^T \pi^1 > 0$. Applying Theorem 1 again, we have $M^{\mathcal{S}_1} x = \mathbf{1}, x > 0$ does not have a solution. Thus, at least two assortments in \mathcal{S}_1 have different total attractiveness. Without loss of generality, we assume A_1 and A_2 in \mathcal{S}_1 have different total attractivenesses. Substituting (11) into (9), we have

$$\lambda \frac{\beta(\mathcal{S}_1)V'(A_n)}{\beta(\mathcal{S}_1)V'(A_n) + 1} = \lambda' \frac{V'(A_n)}{V'(A_n) + 1}, \quad n = 1, 2. \quad (12)$$

Dividing the equation of $n = 1$ with that of $n = 2$, we have

$$\frac{\beta(\mathcal{S}_1)V'(A_1) + 1}{\beta(\mathcal{S}_1)V'(A_2) + 1} = \frac{V'(A_1) + 1}{V'(A_2) + 1}. \quad (13)$$

Since $V'(A_1) \neq V'(A_2)$ and both are positive, we have $\beta(\mathcal{S}_1) = 1$. Thus, $V_j = V'_j$ for $j \in \cup_{A_n \in \mathcal{S}_1} A_n$. Substituting $\beta(\mathcal{S}_1) = 1$ into (9) with n such that $A_n \in \mathcal{S}_1$, we have $\lambda = \lambda'$.

Substituting $\lambda = \lambda'$ into (8) with $A_n \in \mathcal{S}_k$ for any k , we have

$$\frac{\beta(\mathcal{S}_k)}{\beta(\mathcal{S}_k)V'(A_n) + 1} = \frac{1}{V'(A_n) + 1} \Rightarrow \beta(\mathcal{S}_k) = 1.$$

Thus, $V'_j = V_j$ for any $j \in \cup_{A_n \in \mathcal{S}_k} A_n$ with any k .

2. When π satisfies (A.2),

$$\mathbf{1}^T \pi = 0, \quad (M^{\mathcal{S}})^T \pi \leq 0, \quad (M^{\mathcal{S}})^T \pi \neq 0.$$

Similar as the case when π satisfies (A.1), we partition \mathcal{S} into maximal communicating collections $\mathcal{S}_1, \dots, \mathcal{S}_K$. Let π^k denote the portion of π corresponding to \mathcal{S}_k . We have $(M^{\mathcal{S}_k})^T \pi^k \leq 0$ for all k . If for some k_0 , $\mathbf{1}^T \pi^{k_0} > 0$, then for this \mathcal{S}^{k_0} , (A.1) is satisfied and we can then find two assortments in \mathcal{S}^{k_0} with

different total attractivenesses and the argument goes as previous. If for every k , $\mathbf{1}^T \pi^k \leq 0$, then we must have $\mathbf{1}^T \pi^k = 0$ for every k since $\mathbf{1}^T \pi = 0$. Since $(M^S)^T \pi \leq 0$ and $(M^S)^T \pi \neq 0$, for at least one k_0 , $(M^{S_{k_0}})^T \pi^{k_0} \neq 0$. Applying Theorem 1 again, we can find two assortments in S^{k_0} with different total attractivenesses. The argument can go as previous again.

Thus, we have $(V, \lambda) = (V', \lambda')$ and the theorem is proved. \square

In practice, the offer sets made available by the airlines or hotels are often nested by revenue. The companies usually adjust the available assortments by adding a cheaper product or by removing the cheapest available product. That means, for these industries, we will observe some historical offer sets A_1 and A_2 such that A_1 is a strict subset of A_2 . According to Theorem 2, such historical data means the parameters are identifiable.

2.3.2 Piecewise constant arrival rate

In practice, often the case, the arrival rate is non-homogeneous. We now discuss the estimation of the arrival rate *function* and the choice model parameters using the censored sales data and the availability data. We will first focus on the case when the non-homogeneous arrival rate is piecewise constant.

Suppose the arrival rate function is piecewise constant and has M pieces. The arrival rate for the m^{th} piece is λ_m . A collection of N_m distinct historical offer sets, denoted as \mathcal{T}_m , were offered under the m^{th} piece. Let A_{mn} denote the n^{th} offer set in \mathcal{T}_m . Here we also assume that A_{mn} is nonempty. The length of time when A_{mn} is offered is denoted as L_{mn} . There were C_{mnj} purchases of product j with $j \in A_{mn}$ when A_{mn} was offered. The log-likelihood function is as below.

$$\ell\ell(C|\lambda, V) = \sum_{m=1}^M \sum_{n=1}^{N_m} \sum_{j \in A_{mn}} \left(\frac{-\lambda_m L_{mn} V_j}{V(A_{mn}) + 1} + C_{mnj} \log \left(\frac{\lambda_m L_{mn} V_j}{V(A_{mn}) + 1} \right) - \log(C_{mnj}!) \right) \quad (14)$$

Since we assume the choice model parameters are constant throughout the horizon, we do not need each \mathcal{T}_m satisfy the identifiability condition in Theorem 2. As a simple example, suppose the piecewise constant arrival rate function is composed of two pieces. We observe collections of offer sets $\mathcal{T}_1 = \{\{1, 2\}, \{1\}\}$ and $\mathcal{T}_2 = \{\{1\}\}$ for the two pieces, respectively. In this example, \mathcal{T}_1 satisfies the condition in Theorem 2 and we are able to identify λ_1 , V_1 , and V_2 from the sales information using the sales information from the first piece. Once we identify V_2 , we can then use the sales information of product 2 from the second piece to identify λ_2 . The identifiability condition is generalized in Theorem 3 as below. Before stating the theorem, we first introduce the concepts of communicating collections of offer sets.

Definition 4. *Given collections of assortments $\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_m$, \mathcal{T}_1 and \mathcal{T}_2 are said to be communicating on $\mathcal{T}_1, \dots, \mathcal{T}_m$ if there are $A_1 \in \mathcal{T}_1$ and $A_2 \in \mathcal{T}_2$ such that A_1 and A_2 are communicating on $\mathcal{T}_1 \cup \dots \cup \mathcal{T}_m$.*

Definition 5. *Given collections of assortments $\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_m$, these collections are said to be communicating on themselves if every two collections are communicating on $\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_m$.*

We partition the collections of assortments under different arrival rate pieces $\mathcal{T}_1, \dots, \mathcal{T}_M$ into $\mathcal{Q}_1, \dots, \mathcal{Q}_Q$ such that the collections of assortments in each \mathcal{Q}_q are communicating on themselves, and $\mathcal{Q}_q \cup \{\mathcal{T}_m\}$ is not communicating on $\mathcal{Q}_q \cup \{\mathcal{T}_m\}$ for any $\mathcal{T}_m \in \mathcal{Q}_{q'}$ with $q' \neq q$. Thus, each \mathcal{Q}_q is a maximal set of collections of assortments that are communicating.

Note that not every two assortments in \mathcal{Q}_q are communicating on \mathcal{Q}_q . For example, if $\mathcal{T}_1 = \{\{1, 2\}, \{4\}\}$, $\mathcal{T}_2 = \{\{2\}\}$, and $\mathcal{Q}_1 = \{\mathcal{T}_1, \mathcal{T}_2\}$, then \mathcal{T}_1 and \mathcal{T}_2 are communicating but $\{4\}$ is not communicating with other assortments in \mathcal{Q}_q .

Given \mathcal{Q}_q , we further partition all assortments in \mathcal{Q}_q into maximal communicating collections of assortments $\mathcal{S}_{q1}, \dots, \mathcal{S}_{qK_q}$.

The necessary and sufficient condition for the identifiability of the piecewise constant arrival rate function and the choice model parameters are as below.

Theorem 3. *Suppose the customer arrival process is a non-homogeneous Poisson process with piecewise constant rate function. The number and the locations of the break points of the rate function are known. Customers' behavior is governed by a MNL model. Given the sales information C_{mnj} and the historical availability information A_{mn} , the arrival rate function and the choice model parameters are identifiable if and only if for each \mathcal{Q}_q , the following system does not have a solution:*

$$\begin{aligned} M^{\mathcal{T}_m \cap \mathcal{S}_k} x &= \mathbf{1} b_{mk}, & \forall m, k; \\ \frac{b_{mk} d_k}{b_{mk} d_k + 1} e_m &= \frac{b_{mk}}{b_{mk} + 1}, & \forall m, k; \\ x > 0, b_{mk} > 0, d_k > 0, & \forall m, k; \\ \text{not all } e_m &= 1 \end{aligned} \tag{15}$$

Proof. We first show the necessity of the conditions. Suppose for some \mathcal{Q}_q , system (15) has a solution (x^0, b^0, e^0, d^0) . Set $V = x^0$ for products in \mathcal{Q}_q and set the arrival rate $\lambda_m = 1$ for each interval in \mathcal{Q}_q . Since x^0 is a solution to (15), we have $V(A) = b_{mk}^0$ for every $A \in \mathcal{T}_m \cap \mathcal{S}_k$. Scale the attractivenesses of products in $\mathcal{T}_m \cap \mathcal{S}_k$ by d_k . Given this, if we scale λ_m by factor e_m , due to the second set of equations in (15), we know the purchasing rate for every $A \in \mathcal{T}_m \cap \mathcal{S}_k$ will be the same as before for every m and k . Since we always scale the attractivenesses of products in \mathcal{S}_k by the same factor d_k , the relative attractivenesses among them will not be affected. Since not all $e_m = 1$, some arrival rates will be different after the scaling. Thus, we find two sets of parameters with which likelihood functions are the same for all purchase data. Thus, the estimation problem is unidentifiable.

We now prove the sufficiency of the conditions. To achieve this, we show that if the estimation problem is unidentifiable, (15) has a solution for some \mathcal{Q}_q . Given the estimation problem is unidentifiable, there exists $(\lambda, V) \neq (\lambda', V')$ that give rise to

the same likelihood for all purchase data. By equalizing the parameters before C_{mnj} on both sides, we have

$$\frac{\lambda_m V_j}{V(A_{mn}) + 1} = \frac{\lambda'_m V'_j}{V'(A_{mn}) + 1}, \quad \forall m, n, \text{ and } j \in A_{mn}. \quad (16)$$

If products j and j' are in the same assortment, and product j' and l are in another same assortment, then $\frac{V_j}{V'_j} = \frac{V_{j'}}{V'_l} = \frac{V_l}{V'_l}$. In this logic, there exists β_k

$$\frac{V'_j}{V_j} = \beta_k, \quad \forall j \in \bigcup_{A \in \mathcal{S}_k} A, \quad \forall \mathcal{S}_k \in \mathcal{Q}_q. \quad (17)$$

For every assortment A in $\mathcal{T}_m \cap \mathcal{S}_k$, we also have

$$\lambda'_m \frac{\beta_k V(A_{mk})}{\beta_k V(A_{mk}) + 1} = \lambda_m \frac{V(A_{mk})}{V(A_{mk}) + 1}. \quad (18)$$

Without loss of generality, suppose $\lambda_1 \neq \lambda'_1$, then $\beta_k \neq 1$ for all \mathcal{S}_k such that $\mathcal{T}_1 \cap \mathcal{S}_k \neq \emptyset$.

Let A_1 and A_2 be any two assortment in $\mathcal{T}_1 \cap \mathcal{S}_k$ for some k , we have

$$\frac{\lambda'_1 \beta_1 V(A)}{\beta_1 V(A) + 1} = \frac{\lambda_1 V(A)}{V(A) + 1}, \quad \text{for } A = A_1, A_2. \quad (19)$$

Dividing the two equations, we have $\frac{\beta_1 V(A_2) + 1}{\beta_1 V(A_1) + 1} = \frac{V(A_2) + 1}{V(A_1) + 1}$. Since $\beta_1 \neq 1$, we have $V(A_1) = V(A_2)$. In this way, we can show that the assortments in \mathcal{T}_1 belonging to the same communicating collection \mathcal{S}_k for some k have the same total attractivenesses.

Now pick another \mathcal{T}_2 that belongs to the same \mathcal{Q}_q as \mathcal{T}_1 and shares some products in common with \mathcal{T}_1 . Clearly, the shared products have different V and V' . Thus, $\lambda_2 \neq \lambda'_2$. By the same logic as for \mathcal{T}_1 , we can show assortments in $\mathcal{T}_2 \cap \mathcal{S}_k$ for some k have the same total attractiveness. Keep including other \mathcal{T} in \mathcal{Q}_q that shares some products with the \mathcal{T} which has already been discussed. We will have $V(A_1) = V(A_2)$ for any $A_1, A_2 \in \mathcal{T}_m \cap \mathcal{S}_k$ for any $\mathcal{T}_m, \mathcal{S}_k \in \mathcal{Q}_q$.

By setting $x = V$, b_{mk} equal to the common total attractiveness of assortments in the same $\mathcal{T}_m \cap \mathcal{S}_k$, $d_k = \beta_k$, and $e_m = \lambda'/\lambda$, we obtain a solution to (15).

□

If we convert the not-all-equal-to-1 requirement of (15) to $\sum_m (e_m - 1)^2 > 0$, the identifiability conditions (15) can be checked by solving a nonlinear feasibility problem. Also, in practice, strictly nested assortments will probably be offered within the same interval due to the say a revenue management system adjusts the availability. In this case, it is easy to check that the estimation problem is indeed identifiable.

2.3.3 Any integrable arrival rate

When the product attractivenesses are constant and the arrival rate is constant or piecewise constant, the identifiability of the arrival rate and the choice model parameters relies on one essential fact: under at least one interval over which the arrival rate is constant, shall we have two assortments with different total attractivenesses. When this condition is satisfied, with the variation in the total attractiveness and the change in the total purchase rate, we can infer the arrival rate and attractiveness parameters. When the arrival rate is any integrable nonnegative function, even we observe two different assortments that will have different total attractivenesses, we can not attribute the difference between the purchase patterns to the different assortments. The difference may be caused by the change of the arrival rate. Thus, when the structure of the arrival rate function is unknown, one historical sample path is not sufficient for the identification of the arrival rate and choice model parameters.

In this section, we assume we have $H \geq 2$ sample paths of historical availability and purchase information.¹ The h^{th} sample offered nonempty assortment $A_{h,n}$ as its n^{th} assortment. N_h consecutively different assortments were offered in sample h . We observe the historical availability information, the product chosen by the customers, and the occurrence time of the purchases². The total number of purchases in sample

¹When the arrival rate is constant or when the arrival rate is piecewise constant and the number and the locations of the break points are known, we can concatenate the intervals with the same arrival rate from different sample paths into one path. By doing so, the likelihood function does not change.

²In practice, a company most likely will record the transaction times for every traction. When the arrival rate is constant or when the arrival rate is piecewise constant and the number and the

h is I_h . The i^{th} purchase in sample h occurred at time $t_{h,i}$. The horizon length is L . We also let $t_{h,0} = 0$ and $t_{h,I_h+1} = L$. The choice set at time t of sample h is $A_h(t)$. The i^{th} purchase in sample h selected product $c_{h,i}$.

We assume the attractiveness of product j is V_j and it is constant throughout the horizon. The attractiveness of no-purchase is normalized to 1. We assume the Poisson arrival process is non-homogeneous. Given arrival rate function $\lambda(t)$, we define

$$\Lambda_h = \int_0^L \lambda(t) \frac{\sum_{j \in A_h(t)} V_j}{\sum_{j \in A_h(t)} V_j + 1} dt \quad (20)$$

as the accumulative purchase arrival rate during history h . Given this notation, the probability density there is no purchase during time $[0, L] \setminus \{t_{h,i}\}_{i=1, \dots, I_h}$ of history h is $\exp(-\Lambda_h)$. The probability density that in history h , purchases indeed happened at time points $\{t_{h,i}\}$ with purchased products $c_{h,i}$ for $i = 1, \dots, I_h$ is

$$\prod_{i=1}^{I_h} \lambda(t_{h,i}) \frac{V_{c_{h,i}}}{\sum_{j \in A_h(t_{h,i})} V_j + 1}.$$

Thus, the likelihood function of the sales pattern is as Equation (21).

$$\ell(t, c | \lambda, V) = \prod_{h=1}^H \left[\exp(-\Lambda_h) \left(\prod_{i=1}^{I_h} \lambda(t_{h,i}) \frac{V_{c_{h,i}}}{\sum_{j \in A_h(t_{h,i})} V_j + 1} \right) \right]. \quad (21)$$

Given any two arrival rate functions $\lambda(t)$ and $\lambda'(t)$, suppose $\lambda(t) \neq \lambda'(t)$ if and only if $t \in \mathcal{T}$ where $\mathcal{T} \subset [0, L]$. If \mathcal{T} has measure zero, then $\ell(t, c | \lambda, V) = \ell(t, c | \lambda', V)$ almost surely since the probability that some purchases happened during \mathcal{T} is zero. Thus, the maximum likelihood estimator can at most identify two arrival rate functions if they differ over a nonzero measure set. On the other hand, if two arrival rate functions differ only over a zero measure set, then their cumulative arrival rate functions are identical. Given that in most revenue management models, the cumulative rate matters is the only interesting input information about arrival, such identifiability should be sufficient. Thus, in the subsequent discussion, the identifiability of

locations of the break points are known, this detailed transaction time information does not provide extra information than the aggregated sales information for the purpose of estimation.

the arrival rate function will be in this sense, and the equality between arrival rate functions means they are equal almost everywhere.

Let $\mathcal{D}(t) := \{A_h(t) : h = 1, \dots, H\}$ be the collection of assortments offered across different samples at time t . In each sample, a finite number of consecutively different assortments were offered, each assortment is offered for certain positive amount of time. Thus there are a finite number of distinctive $\mathcal{D}(t)$ values throughout the horizon. We represent them as $\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_M$.

We partition $\{\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_M\}$ into $\mathcal{Z}_1, \dots, \mathcal{Z}_Z$ such that the collections in each \mathcal{Z}_z are communicating on themselves (as defined in Definition 5), and $\mathcal{Z}_z \cup \{\mathcal{D}_m\}$ is not communicating on $\mathcal{Z}_z \cup \{\mathcal{D}_m\}$ with any $\mathcal{D}_m \notin \mathcal{Z}_z$. Thus, each \mathcal{Z}_z is a maximal set of collections of assortments that is communicating. The sufficient and necessary conditions for the identifiability of the piecewise constant rate function $\lambda(t)$ and the choice parameter V are stated as below. Given \mathcal{Z}_m , we partition the assortments in it into maximal communicating collections of assortments $\mathcal{S}_{z1}, \dots, \mathcal{S}_{zk_m}$.

Theorem 4. *Suppose the Poisson arrival rate $\lambda(t)$ is an integrable nonnegative function. $\lambda(t)$ and the choice parameters V are identifiable if and only if for each \mathcal{Z}_m the following system does not have a solution:*

$$\begin{aligned} M^{\mathcal{D}_m \cap \mathcal{S}_k} x &= \mathbf{1} b_{mk}, & \forall m, k; \\ \frac{b_{mk} d_k}{b_{mk} d_k + 1} e_m &= \frac{b_{mk}}{b_{mk} + 1}, & \forall m, k; \\ x > 0, b_{mk} > 0, d_k > 0, & \forall m, k; \\ \text{not all } e_m &= 1 \end{aligned} \tag{22}$$

The following lemma states the necessary condition for the equality of the likelihood functions of two different sets of parameters $(\lambda(t), V)$ and $(\lambda'(t), V')$. The lemma will be useful in the proof of the identifiability conditions.

Lemma 1. *Given $(\lambda(t), V) \neq (\lambda'(t), V)$, if $\ell(t, c|\lambda, V) = \ell(t, c|\lambda', V')$ for almost all*

purchase patterns (t, c) , then

$$\lambda(t) \frac{V_c}{\sum_{j \in A_h(t_{h,i})} V_j + 1} = \lambda'(t) \frac{V'_c}{\sum_{j \in A_h(t_{h,i})} V'_j + 1} \quad (23)$$

for every c and almost every t .

Proof. We prove the result by contradiction. Suppose for some h_0 and c , there exists a set \mathcal{T} of t where Equation (23) is violated and the measure of \mathcal{T} is positive. We further partition \mathcal{T} into \mathcal{T}^+ and \mathcal{T}^- as below

$$\mathcal{T}^+ = \left\{ t \in \mathcal{T} : \lambda(t) \frac{V_c}{\sum_{j \in A_h(t_{h,i})} V_j + 1} > \lambda'(t) \frac{V'_c}{\sum_{j \in A_h(t_{h,i})} V'_j + 1} \right\}, \quad (24)$$

and

$$\mathcal{T}^- = \left\{ t \in \mathcal{T} : \lambda(t) \frac{V_c}{\sum_{j \in A_h(t_{h,i})} V_j + 1} < \lambda'(t) \frac{V'_c}{\sum_{j \in A_h(t_{h,i})} V'_j + 1} \right\}.$$

Without loss of generality, we assume \mathcal{T}^+ has positive measure. Since $\ell(t, c|\lambda, V) = \ell(t, c|\lambda', V')$ for almost all purchase patterns (t, c) , we first select (t_0, c_0) corresponding to the case where no purchase happened during any histories. Then the equality between the likelihood functions implies

$$\exp \left(\sum_{h=1}^H \Lambda_h \right) = \exp \left(\sum_{h=1}^H \Lambda'_h \right). \quad (25)$$

We further select purchase pattern (t_+, c_+) corresponding to the case where one purchase happened at t^+ of \mathcal{T}^+ of history h_0 and no other purchases happened. Note such purchase pattern has positive probability since \mathcal{T}^+ has positive measure. Thus, $\ell(t, c|\lambda, V) = \ell(t, c|\lambda', V')$ holds for (t_+, c_+) implies

$$\exp \left(\sum_{h=1}^H \Lambda_h \right) \lambda(t_+) \frac{V_{c_+}}{\sum_{j \in A_h(t_+)} V_j + 1} = \exp \left(\sum_{h=1}^H \Lambda'_h \right) \lambda'(t_+) \frac{V'_{c_+}}{\sum_{j \in A_h(t_+)} V'_j + 1}. \quad (26)$$

(25) and (26) imply

$$\lambda(t_+) \frac{V_{c_+}}{\sum_{j \in A_h(t_+)} V_j + 1} = \lambda'(t_+) \frac{V'_{c_+}}{\sum_{j \in A_h(t_+)} V'_j + 1},$$

which contradicts the facts in (24). Thus, the result is established. \square

Now we are ready to prove Theorem 4.

Proof of Theorem 4. We first show the necessity. Suppose system (22) has a solution (\bar{x}, b, d, e) for one z . Here we let \bar{x} denote the portion of products corresponding to the products that were offered in \mathcal{Z}_z . Set the product attractivenesses in \mathcal{Z}_z equal to \bar{x} . According to the first equation of (22), $V(A_{mk}) = b_{mk}$ for any $A_{mk} \in \mathcal{D}_m \cap \mathcal{S}_k$. Scale the attractivenesses of products in \mathcal{S}_k by d_k . Scale the arrival rate corresponding to \mathcal{D}_m by e_m . According to the second equation of (22), the purchase rate before and after scaling are the same. Since we scale the product attractivenesses in the same \mathcal{S}_k with the same scaler, their relative attractivenesses will stay the same. Since not all $e_m = 1$, we find two different set of parameters and their corresponding likelihood function are identical. Thus, the estimation problem is unidentifiable.

We now prove the sufficiency. Suppose $(\lambda(t), V)$ and $(\lambda'(t), V')$ have the same likelihood function. According to Lemma 1, the equality between the likelihood functions means there exists $\mathcal{T} \subset [0, L]$ such that the measure of \mathcal{T} is L and for any purchase patterns (t, c) , Equation (23) holds for $t \in \mathcal{T}$. For $m = 1, \dots, M$, let t_m be any time point in \mathcal{T} such that $\mathcal{D}(t) = \mathcal{D}_m$. Using Lemma 1, we then have

$$\lambda(t_m) \frac{V_j}{\sum_{j \in A_h(t_m)} V_j + 1} = \lambda'(t_m) \frac{V'_j}{\sum_{j \in A_h(t_m)} V'_j + 1}, \forall m, h, \text{ and } j \in A_h(t_m). \quad (27)$$

According to (16), (27) can be viewed as the implication of the equality of the likelihood function in the case where

1. the arrival rate is piecewise constant with known break points, and
2. under the m^{th} piece, the offered assortments are \mathcal{D}_m .

According to the Theorem 3, (27) and the unsolvability of (22) implies $V = V'$ and $\lambda(t_m) = \lambda'(t_m)$ for $m = 1, \dots, M$. Since t_m can be almost every point in $[0, L]$, we have $\lambda(t) = \lambda'(t)$ for almost every t . Thus, the sufficiency is proved.

□

Remark 1. *The identifiability of the piecewise constant arrival rate function and the product attractiveness does not depend on the number of break points in the arrival rate function. Knowing that every function could be approximated by a piecewise linear function, we expect when the assortments from the historical samples satisfies the conditions in Theorem 4, we could approximately recover the true arrival rate function (not necessary piecewise constant) and the product attractiveness. On the other hand, given any number of historical data, if we allow the decision space to include all piecewise constant function when maximizing the likelihood function (21), the optimal solution will try to put all the arrival intensity to the time points where a purchased occurred, with zero arrival rate at other intervals. We should carefully select the decision space to avoid such nonsense solutions.*

2.4 Estimation Method

We now propose an estimation method to estimate the arrival rate function $\lambda(t)$ and the choice model parameters V . The available data for estimation are multiple historical samples of the availability and the purchase information. Specifically, from each sample, we know the following:

1. historical assortments (we know the assortment used at any time point for any sample path);
2. occurrence time of each purchase;
3. the product chosen at each purchase.

Such data requirement is not restricted in practice. For example, an airline company typically estimates an arrival model and a choice model for each day of week. For each given day of week, the company can extract multiple streams of availability and booking information for the day of the week. Such information can be used in our method to estimate $\lambda(t)$ and V .

Suppose we have $H \geq 2$ historical samples. As the notation used in Section 2.3.3, we let $A_h(t)$ denote the assortment available at time t in sample h . There are I_h purchases in sample h and let $I = \sum_{h=1}^H I_h$. The i^{th} purchase in sample h occurred at time $t_{h,i}$. We merge the occurrence times of all the purchases from all the histories and sort them in ascending order. Let $t_{(i)}$ denote the occurrence time of the i^{th} purchase for $i = 1, \dots, I - 1$. For notational convenience, we define $t_{(0)} = 0$ and $t_{(I)} = L$. Let $c_{(i)}$ denote the product chosen by the i^{th} purchase. Let $A_{(i)}$ denote the assortment available when the i^{th} purchase happened.

We also let B to denote the set of break points of the assortments from all history data. That is, if $t \in B$, then for some history, the assortment got changed at time t . Given B , we can divide the time horizon to $|B| + 1$ subintervals. We use \mathcal{B} to denote the set of these subintervals. According to the definition of \mathcal{B} , for any subinterval $b \in \mathcal{B}$, the assortment during b of any history is constant. We let A_{hb} denote the assortment during subinterval b of history h . Let C_{hbj} denote the number of purchases of product j during subinterval b of history h . Let C_{hb} denote vector of purchases of all products.

Our estimation method has two steps. The first step estimates the product attractivenesses V using a maximum likelihood estimator. Given the estimated V , the second step constructs the cumulative arrival rate function.

2.4.1 Estimation of V

Suppose the product attractivenesses are V and the cumulative arrival rate for interval b is Λ_b . Given these parameters, the likelihood of the purchase pattern during interval b of history h is as following:

$$f(C_{hb}|\Lambda_b, V, A_{hb}) = \frac{\exp\left(-\Lambda_b \frac{V(A_{hb})}{V(A_{hb})+1}\right) \left(\Lambda_b \frac{V(A_{hb})}{V(A_{hb})+1}\right)^{\sum_{j \in A_{hb}} C_{hbj}}}{\prod_{j \in A_{hb}} C_{hbj}!} \prod_{j \in A_{hb}} \left(\frac{V_j}{V(A_{hb})}\right)^{C_{hbj}}$$

The scaled log-likelihood for the whole purchase pattern is as following:

$$LL^H(\Lambda, V) = \frac{1}{H} \sum_{b \in \mathcal{B}} \sum_{h=1}^H \log f(C_{hb} | \Lambda_b, V, A_{hb}) \quad (28)$$

Given any V , $LL^H(\Lambda, V)$ is concave in Λ . According to the first order condition, the optimal Λ is given by

$$\Lambda_b^H(V) = \frac{\sum_{h=1}^H \sum_{j \in A_{hb}} C_{hbj}}{\sum_{h=1}^H \frac{V(A_{hb})}{V(A_{hb})+1}}. \quad (29)$$

Substituting $\Lambda^H(V)$ back to $LL^H(\Lambda, V)$, we can reduce optimization problem as following:

$$\max_{V \in \Theta} : LL^H(\Lambda^H(V), V) = \frac{1}{H} \sum_{b=1}^{|B|+1} \sum_{h=1}^H \log f(C_{hb} | \Lambda_b^H(V), V, A_{hb}) \quad (30)$$

where $\Theta = \{V : \underline{V} \leq V_j \leq \bar{V}, \forall j\}$ and \underline{V} and \bar{V} are known positive quantities. The optimal solution V from (30) is our estimate of the product attractivenesses.

2.4.2 Construction of cumulative arrival rate function

We now describe how to construct the estimated cumulative arrival rate function after we have an estimate V of the product attractivenesses.

For the i^{th} purchase among all purchases, define a_i as below:

$$a_i = \sum_{h=1}^H \int_{t_{(i-1)}}^{t_{(i)}} \frac{\sum_{j \in A_h(t)} V_j}{\sum_{j \in A_h(t)} V_j + 1} dt.$$

The integral in the definition of a_i denotes the effective time length for purchases. It is the natural time length filtered by the purchasing probability. Since for a given history in practice, $A_h(t)$ is a piecewise function, the integral should also be easy to calculate. For example, if the $(i-1)^{\text{st}}$ purchase and the i^{th} are both during subinterval b , then

$$a_i = \sum_{h=1}^H (t_{(i)} - t_{(i-1)}) \frac{V(A_{hb})}{V(A_{hb}) + 1}.$$

Given a_i , define the cumulative arrival rate at $t_{(i)}$ as following:

$$\Lambda(t_{(i)}) = \sum_{i'=1}^i \frac{t_{(i)} - t_{(i-1)}}{a_i}, \quad i = 1, \dots, I. \quad (31)$$

Once we determine the cumulative arrival rate $t_{(i)}$ for all i 's, we connect the consecutive $\Lambda(t_{(i)})$'s using a piecewise linear function. This gives us the estimated cumulative arrival rate function.

2.4.3 Convergence proof

In this section, we prove that the choice model parameters from the estimation model (30) and the cumulative arrival rate function using interpolation points as (31) converge to the true quantities almost surely, as the sample size H increases. Throughout this section, we assume that the number of breakpoints in B is bounded as H increases.

2.4.3.1 Convergence of V

Let H_{bA} denote the set of history samples among H total samples such that $A_{hb} = A$. Let C_{bA}^h denote the vector of the numbers of purchases for all products during interval b of the sample h when $A_{hb} = A$. $\Lambda_b^H(V)$ defined as in (29) can be rewritten as:

$$\Lambda_b^H(V) = \frac{\sum_{h=1}^H \sum_{j \in A_{hb}} C_{hbj}}{\sum_{h=1}^H \frac{\sum_{j \in A_{hb}} V_j}{\sum_{j \in A_{hb}} V_j + 1}} = \frac{\sum_{A \in 2^J} \sum_{h \in H_{bA}} \sum_{j \in A} C_{bAj}^h}{\sum_{A \in 2^J} |H_{bA}| \frac{\sum_{j \in A} V_j}{\sum_{j \in A} V_j + 1}} \quad (32)$$

And the log-likelihood function $LL^H(\Lambda^H(V), V)$ can be rewritten as:

$$\begin{aligned} LL^H(\Lambda^H(V), V) &= \frac{1}{H} \sum_{b=1}^{|B|+1} \sum_{A \in 2^J} \sum_{h \in H_{bA}} \log f(C_{bA}^h | \Lambda_b^H(V), V, A) \\ &= \sum_{b=1}^{|B|+1} \sum_{A \in 2^J} \frac{|H_{bA}|}{H} \frac{1}{|H_{bA}|} \sum_{h \in H_{bA}} \log f(C_{bA}^h | \Lambda_b^H(V), V, A) \end{aligned} \quad (33)$$

For each b and A , we assume that $\lim_{H \rightarrow \infty} \frac{|H_{bA}|}{H}$ exists and we denote the limit as α_{bA} . Let Λ_b^0 denote the true cumulative arrival rate during interval b . Let V_j^0 denote the true attractiveness of product j .

Lemma 2. For given b , if $\lim_{H \rightarrow \infty} \frac{|H_{bA}|}{H} = \alpha_{bA}$ for every A , then $\Lambda_b^H(V)$ on Θ uniformly

converges to $\Lambda_b(V) := \frac{\sum_{A \in 2^J} \alpha_{bA} \Lambda_b^0 \frac{\sum_{j \in A} V_j^0}{\sum_{j \in A} V_j^0 + 1}}{\sum_{A \in 2^J} \alpha_{bA} \frac{\sum_{j \in A} V_j}{\sum_{j \in A} V_j + 1}}$ almost surely as H goes to infinity.

Proof. We can divide both the numerator and the denominator of (32) by H and we get

$$\Lambda_b^H(V) = \frac{\frac{1}{H} \sum_{A \in 2^J} \sum_{h \in H_{bA}} \sum_{j \in J} C_{bAj}^h}{\frac{1}{H} \sum_{A \in 2^J} |H_{bA}| \frac{\sum_{j \in A} V_j}{\sum_{j \in A} V_j + 1}}.$$

For the numerator, we know the limit $\lim_{H \rightarrow \infty} \frac{|H_{bA}|}{H}$ exists. We analyze the term

$$\lim_{H \rightarrow \infty} \frac{1}{H} \sum_{h \in H_{bA}} \sum_{j \in A} C_{bAj}^h$$

based on the boundedness of $|H_{bA}|$.

- When $|H_{bA}|$ is unbounded, according to the strong law of large numbers,

$$\frac{1}{|H_{bA}|} \sum_{h \in H_{bA}} \sum_{j \in A} C_{bAj}^h$$

converges to $\Lambda_b^0 \frac{\sum_{j \in A} V_j^0}{\sum_{j \in A} V_j^0 + 1}$ almost surely as H goes to infinity. Since $\frac{|H_{bA}|}{H}$ converges to α_{bA} deterministically, $\frac{1}{H} \sum_{h \in H_{bA}} \sum_{j \in A} C_{bAj}^h$ converges to $\alpha_{bA} \Lambda_b^0 \frac{\sum_{j \in A} V_j^0}{\sum_{j \in A} V_j^0 + 1}$ almost surely as H goes to infinity.

- When $|H_{bA}|$ is bounded, $\lim_{H \rightarrow \infty} \sum_{h \in H_{bA}} \sum_{j \in A} C_{bAj}^h$ is bounded almost surely.

Thus, $\frac{1}{H} \sum_{h \in H_{bA}} \sum_{j \in A} C_{bAj}^h$ converges to 0 almost surely as H goes to infinity. Because $\alpha_{bA} = 0$ when $|H_{bA}|$ is bounded, we can also state the result as

$\frac{1}{H} \sum_{h \in H_{bA}} \sum_{j \in A} C_{bAj}^h$ converges to $\alpha_{bA} \Lambda_b^0 \frac{\sum_{j \in A} V_j^0}{\sum_{j \in A} V_j^0 + 1}$ almost surely.

Since the numerator is a constant function of V , the convergence is also uniform on Θ .

Thus, $\frac{1}{H} \sum_{A \in 2^J} \sum_{h \in H_{bA}} \sum_{j \in J} C_{bAj}^h$ on Θ uniformly converges to $\sum_{A \in 2^J} \alpha_{bA} \Lambda_b^0 \frac{\sum_{j \in A} V_j^0}{\sum_{j \in A} V_j^0 + 1}$ almost surely.

By the convergence of $\frac{|H_{bA}|}{H}$ and the boundedness of Θ , the denominator

$$\frac{1}{H} \sum_{A \in 2^J} |H_{bA}| \frac{\sum_{j \in A} V_j}{\sum_{j \in A} V_j + 1}$$

also uniformly converges to $\sum_{A \in 2^J} \alpha_{bA} \frac{\sum_{j \in A} V_j}{\sum_{j \in A} V_{j+1}}$ on Θ .

Combining the convergence results of the numerator and the denominator, we prove the results. \square

The following lemma is from Lemma 1 of [52]. It gives the conditions under which a sequence of sample average function uniformly converge to the expected function almost surely as the sample size increases.

Lemma 3. *Suppose a compact set Θ , a random vector Y with dimension $m \times 1$ and distribution function G , and a function $\phi : \mathbb{R}^{m \times 1} \rightarrow \mathbb{R}$ are given. If the following conditions are satisfied:*

1. $\phi(y, \theta)$ is G -measurable for each $\theta \in \Theta$;
2. There exists $b(y)$ such that $|\phi(y, \theta)| \leq b(y)$ for each $\theta \in \Theta$ and $b(y)$ is G -integrable;
3. ϕ is almost surely continuous in the sense that for each fixed $\theta \in \Theta$, the set $\{y : \lim_{\gamma \rightarrow \theta} \phi(y, \gamma) = \phi(y, \theta)\}$ has probability 1.

then $P \left(\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^n \phi(Y_i, \theta) - \int \phi(y, \theta) dG(y) \right| = 0 \right) = 1$.

Let $\Gamma := \{(\Lambda, V) : V \in \Theta, \Lambda \in [\Lambda(V) - \delta, \Lambda(V) + \delta]\}$ for some sufficiently small $\delta > 0$ such that Λ and V in Γ are all bounded away from 0. Then Γ is a compact set.

Lemma 4. *For given b , if $\lim_{H \rightarrow \infty} \frac{|H_{bA}|}{H} = \alpha_{bA}$ for every A , then $\frac{1}{H} \sum_{h \in H_{bA}} \log f(C_{bA}^h | \Lambda_b, V, A)$ on Γ uniformly converges to $\alpha_{bA} E[\log f(C_{bA} | \Lambda_b, V, A)]$ almost surely as H goes to infinity. That is:*

$$P \left(\lim_{H \rightarrow \infty} \sup_{(\Lambda, V) \in \Gamma} \left| \frac{1}{H} \sum_{h \in H_{bA}} \log f(C_{bA}^h | \Lambda_b, V, A) - \alpha_{bA} E[\log f(C_{bA} | \Lambda_b, V, A)] \right| = 0 \right) = 1 \quad (34)$$

Proof. We prove the results based on the boundedness of H_{bA} .

- When H_{bA} is bounded. In this case, $\alpha_{bA} = 0$ and $\sum_{h \in H_{bA}} \log f(C_{bA}^h | \Lambda_b, V, A)$ is the summation of a finitely many random variables. Since Λ and V in Γ are bounded away from 0, $\sup_{(\Lambda, V) \in \Gamma} \sum_{h \in H_{bA}} \log f(C_{bA}^h | \Lambda_b, V, A)$ is finite almost surely. Then

$$P \left(\lim_{H \rightarrow \infty} \sup_{(\Lambda, V) \in \Gamma} \left| \frac{1}{H} \sum_{h \in H_{bA}} \log f(C_{bA}^h | \Lambda_b, V, A) - \alpha_{bA} E[\log f(C_{bA} | \Lambda_b, V, A)] \right| = 0 \right) \quad (35)$$

$$= P \left(\lim_{H \rightarrow \infty} \frac{1}{H} \sup_{(\Lambda, V) \in \Gamma} \left| \sum_{h \in H_{bA}} \log f(C_{bA}^h | \Lambda_b, V, A) \right| = 0 \right) = 1 \quad (36)$$

- When H_{bA} is unbounded. It is easy to check that $\log f(C_{hb} | \Lambda_b, V, A)$ is almost surely continuous on Γ . Since Λ and V in Γ are bounded away from 0,

$$\sup_{(\Lambda, V) \in \Gamma} |\log f(C_{hb}, A | \Lambda, V)| < \infty$$

and is integrable. According to Lemma 3, $\frac{1}{|H_{bA}|} \sum_{h \in H_{bA}} \log f(C_{bA}^h | \Lambda_b, V, A)$ on Γ uniformly converges to $E[\log f(C_{bA} | \Lambda_b, V, A)]$ almost surely. Since $\frac{|H_{bA}|}{H}$ deterministically converges to α_{bA} , $\frac{1}{H} \sum_{h \in H_{bA}} \log f(C_{bA}^h | \Lambda_b, V, A)$ on Γ uniformly converges to $\alpha_{bA} E[\log f(C_{bA} | \Lambda_b, V, A)]$ almost surely

To conclude, we prove the results. □

Lemma 5. *For given b , if $\lim_{H \rightarrow \infty} \frac{|H_{bA}|}{H} = \alpha_{bA}$ for every A , then*

$$\frac{1}{H} \sum_{h \in H_{bA}} \log f(C_{bA}^h | \Lambda_b^H(V), V, A)$$

on Θ uniformly converges to $\alpha_{bA} E[\log f(C_{bA} | \Lambda_b(V), V, A)]$ almost surely as H goes to infinity. That is, for any given $\varepsilon > 0$,

$$P \left(\lim_{H \rightarrow \infty} \sup_{V \in \Theta} \left| \frac{1}{H} \sum_{h \in H_{bA}} \log f(C_{bA}^h | \Lambda_b^H(V), V, A) - \alpha_{bA} E[\log f(C_{bA} | \Lambda_b(V), V, A)] \right| > \varepsilon \right) = 0 \quad (37)$$

Proof. Given $\varepsilon > 0$,

$$\begin{aligned}
& P \left(\lim_{H \rightarrow \infty} \sup_{V \in \Theta} \left| \frac{1}{H} \sum_{h \in H_{bA}} \log f(C_{bA}^h | \Lambda_b^H(V), V, A) - \alpha_{bA} E[\log f(C_{bA} | \Lambda_b(V), V, A)] \right| > \varepsilon \right) \\
& \leq P \left(\lim_{H \rightarrow \infty} \sup_{V \in \Theta} \left| \frac{1}{H} \sum_{h \in H_{bA}} \log f(C_{bA}^h | \Lambda_b^H(V), V, A) - \alpha_{bA} E[\log f(C_{bA} | \Lambda_b^H(V), V, A)] \right| > \frac{\varepsilon}{2} \right) \\
& \quad (38)
\end{aligned}$$

$$\begin{aligned}
& + P \left(\lim_{H \rightarrow \infty} \sup_{V \in \Theta} |\alpha_{bA} E[\log f(C_{bA} | \Lambda_b^H(V), V, A)] - \alpha_{bA} E[\log f(C_{bA} | \Lambda_b(V), V, A)]| > \frac{\varepsilon}{2} \right) \\
& \quad (39)
\end{aligned}$$

We now consider the two terms in turn. For the first term, since $\Lambda^H(V)$ on Θ uniformly converges to $\Lambda(V)$ almost surely, when H is large enough (H could depend on sample path), $(\Lambda^H(V), V) \in \Gamma$ for $V \in \Theta$. By Lemma 4,

$$\begin{aligned}
& P \left(\lim_{H \rightarrow \infty} \sup_{V \in \Theta} \left| \frac{1}{H} \sum_{h \in H_{bA}} \log f(C_{bA}^h | \Lambda_b^H(V), V, A) - \alpha_{bA} E[\log f(C_{bA} | \Lambda_b^H(V), V, A)] \right| > \frac{\varepsilon}{2} \right) \\
& = P \left(\lim_{H \rightarrow \infty} \sup_{(\Lambda, V) \in \Gamma} \left| \frac{1}{H} \sum_{h \in H_{bA}} \log f(C_{bA}^h | \Lambda_b, V, A) - \alpha_{bA} E[\log f(C_{bA} | \Lambda_b, V, A)] \right| > \frac{\varepsilon}{2} \right) = 0
\end{aligned}$$

For the second term, since $E[\log f(C_{bA} | \Lambda_b, V, A)]$ is Lipschitz continuous on Γ , there exists δ^* such that if $|\Lambda_b - \Lambda'_b| \leq \delta^*$,

$$\alpha_{bA} \sup_{V \in \Theta} |E[\log f(C_{bA} | \Lambda_b, V, A)] - E[\log f(C_{bA} | \Lambda'_b, V, A)]| \leq \frac{\varepsilon}{2}.$$

Since $\Lambda_b^H(V)$ on Θ uniformly converges to $\Lambda_b(V)$ almost surely, for almost every sample path, when H is sufficiently large, $\sup_{V \in \Theta} |\Lambda_b^H(V) - \Lambda_b(V)| \leq \delta^*$. Thus, the second term is also equal to 0.

In summary, we have for every $\varepsilon > 0$,

$$P \left(\lim_{H \rightarrow \infty} \sup_{V \in \Theta} \left| \frac{1}{H} \sum_{h \in H_{bA}} \log f(C_{bA}^h | \Lambda_b^H(V), V, A) - \alpha_{bA} E[\log f(C_{bA} | \Lambda_b(V), V, A)] \right| > \varepsilon \right) = 0$$

and the lemma is proved. \square

We can now apply Lemma 5 to the objective function of the auxiliary estimator.

Corollary 1. Suppose for every b and every A , $\lim_{H \rightarrow \infty} \frac{|H_{bA}|}{H} = \alpha_{bA}$, then $LL^H(\Lambda^H(V), V)$ on Θ uniformly converges to $\sum_{b=1}^{|B|+1} \sum_{A \in S_b^+} \alpha_{bA} \mathbb{E}[\log f(C_{hb}|\Lambda(V), V, A)]$ almost surely. That is:

$$P \left(\lim_{H \rightarrow \infty} \sup_{V \in \Theta} \left| LL^H(\Lambda^H(V), V) - \sum_{b=1}^{|B|+1} \sum_{A \in S_b^+} \alpha_{bA} \mathbb{E}[\log f(C_{hb}|\Lambda(V), V, A)] \right| = 0 \right) = 1$$

Let S_b^+ denote the collection of assortments such that $\alpha_{bA} > 0$ if $A \in S_b^+$ and $\alpha_{bA} = 0$ if $A \notin S_b^+$. The following result is a direct application of the information inequality.

Lemma 6. If $\{S_b^+ : b = 1, \dots, |B| + 1\}$ satisfies the identifiability condition in Theorem 4, then $(\Lambda(V^0), V^0)$ is the unique maximizer of the following problem:

$$\max_{V \in \Theta} \sum_{b=1}^{|B|+1} \sum_{A \in S_b^+} \alpha_{bA} \mathbb{E} \log f(C_{bA}|\Lambda_b(V), V, A)$$

Lemma 7. Suppose there is a function $Q_0(\theta)$ such that a) $Q_0(\theta)$ is uniquely maximized at θ_0 ; b) Θ is compact; c) $Q_0(\theta)$ is continuous; d) $Q_n(\theta)$ converges to $Q_0(\theta)$ in the sense that

$$P \left(\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta} |Q_n(\theta) - Q_0(\theta)| = 0 \right) = 1.$$

If θ_n is a maximizer of $Q_n(\theta)$ within Θ , then θ_n converges to θ_0 almost surely.

Proof. For any $\varepsilon > 0$, since θ_n is a maximizer of Q_n , $Q_n(\theta_n) > Q_n(\theta_0) - \frac{\varepsilon}{3}$. Because of the convergence condition d),

$$\begin{aligned} P \left(\lim_{n \rightarrow \infty} (Q_0(\theta_n) - Q_n(\theta_n)) > -\frac{\varepsilon}{3} \right) &= 1 - P \left(\lim_{n \rightarrow \infty} (Q_0(\theta_n) - Q_n(\theta_n)) \leq -\frac{\varepsilon}{3} \right) \\ &\geq 1 - P \left(\lim_{n \rightarrow \infty} |Q_0(\theta_n) - Q_n(\theta_n)| \geq \frac{\varepsilon}{3} \right) \\ &\geq 1 - P \left(\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta} |Q_0(\theta) - Q_n(\theta)| \geq \frac{\varepsilon}{3} \right) \\ &= 1 \end{aligned}$$

Similarly, we have $Q_n(\theta_0) > Q_0(\theta_0) - \frac{\varepsilon}{3}$ with probability approaching 1 as n increases. Combining these three inequalities, we have

$$Q_0(\theta_n) > Q_n(\theta_n) - \frac{\varepsilon}{3} > Q_0(\theta_0) - \frac{2\varepsilon}{3} > Q_0(\theta_0) - \varepsilon$$

with probability 1. Thus, for any $\varepsilon > 0$, $Q_0(\theta_n) > Q_0(\theta_0) - \varepsilon$ with probability approaching 1 as n increases. Let \mathcal{N} be any open subset of Θ around θ_0 . Clearly $\Theta \cap \mathcal{N}^c$ is compact. Since $Q_0(\theta)$ is continuous and is uniquely maximized at θ_0 , we have $\sup_{\theta \in \Theta \cap \mathcal{N}^c} Q_0(\theta) < Q_0(\theta_0)$. Choosing $\varepsilon = Q_0(\theta_0) - \sup_{\theta \in \Theta \cap \mathcal{N}^c} Q_0(\theta)$, it follows that $Q_0(\theta_n) > \sup_{\theta \in \Theta \cap \mathcal{N}^c} Q_0(\theta)$ with probability approaching 1 as n increases. Thus, $\theta_n \in \mathcal{N}$ with probability approaching 1 as n increases. We can choose the neighborhood arbitrarily small and the result is proved. \square

Theorem 5. *If $\lim_{H \rightarrow \infty} \frac{|H_{bA}|}{H} = \alpha_{bA}$ for every A and b , and the set of collections of assortments in $\{S_b^+ : b = 1, \dots, |B| + 1\}$ satisfies the identifiability condition in Theorem 4, then the estimate of V from (30) converges to V^0 almost surely as H goes to infinity.*

Proof. According to Corollary 1, the sample log-likelihood function uniformly converges to its expected version almost surely as sample size increases. According to Lemma 6, when the identifiability condition is satisfied, the expected log-likelihood function has a unique maximizer. By applying Lemma 7, we have the almost sure convergence of the estimate of V . \square

2.4.3.2 Convergence of $\Lambda(t)$

We now prove the convergence of the cumulative arrival rate function. We first show that the contribution from a single purchase to the cumulative arrival rate is negligible. Let $l_i = t_{(i)} - t_{(i-1)}$.

Lemma 8. $\sup_i \sup_{V \in \Theta} \frac{l_i}{a_i} = 0$ almost surely as H goes to infinity.

Proof. Let l_{ib} denote the length of $b \cup [t_{i-1}, t_i]$. Then:

$$\frac{l_i}{a_i} = \frac{\sum_{b \in \mathcal{B}} l_{ib}}{\sum_{b \in \mathcal{B}} l_{ib} \sum_{h=1}^H \frac{V(A_{bh})}{V(A_{bh})+1}}$$

Since each A_{hb} is nonempty and $V \in \Theta$, $\frac{V(A_{hb})}{V(A_{hb})+1} \geq \frac{\underline{V}}{\underline{V}+1}$ for any h and b . And we have

$$\sup_i \sup_{V \in \Theta} \frac{\sum_{b \in \mathcal{B}} l_{ib}}{\sum_{b \in \mathcal{B}} l_{ib} \sum_{h=1}^H \frac{V(A_{bh})}{V(A_{bh})+1}} \leq \sup_i \frac{\sum_{b \in \mathcal{B}} l_{ib}}{\sum_{b \in \mathcal{B}} l_{ib} H \frac{\underline{V}}{\underline{V}+1}} = \frac{\underline{V}+1}{H\underline{V}}$$

The right hand side goes to 0 almost surely as H goes to infinity. Thus, the result is proved. \square

For any $t \in (0, L]$, let I_t denote the number of purchases from all histories before time t .

Lemma 9. *For any $t \in (0, L]$, let $b(t)$ denote the intersection between interval b and interval $[0, t]$. Let $\Lambda^0(b(t))$ denote the true cumulative arrival rate for the interval $b(t)$. Suppose for any b and A , $\lim_H \frac{H_b(A)}{H} = \alpha_{bA}$. If $V \in \Theta$, then $\sum_{i=1}^{I_t} \frac{l_i}{a_i}$ on Θ uniformly converges to the following almost surely as H goes to infinity:*

$$\sum_b \frac{\Lambda^0(b(t)) \sum_A \alpha_{bA} \frac{V^0(A)}{V^0(A)+1}}{\sum_A \alpha_{bA} \frac{V(A)}{V(A)+1}}.$$

Proof. For each given interval $b(t)$, except for the first purchase during this interval whose previous purchase is in interval $b-1$, for all other purchases,

$$\frac{l_i}{a_i} = \frac{1}{\sum_{h=1}^H \frac{V(A_{bh})}{V(A_{bh})+1}}.$$

Thus,

$$\sum_{i:t_i \in b(t)} \frac{l_i}{a_i} = \frac{\left(\sum_{i:t_i \in b(t)} 1\right) - 1}{\sum_{h=1}^H \frac{V(A_{bh})}{V(A_{bh})+1}} + \frac{l'_b}{a'_b}, \quad (40)$$

where $\frac{l'_b}{a'_b}$ denote the contribution from the first purchase during $b(t)$. According to Lemma 8, $\frac{l'_b}{a'_b}$ on Θ uniformly converges to 0 almost surely. Since $|B|$ is bounded, the

sum of such terms from all intervals also uniformly converges to 0 almost surely. For the first term on the right hand side of (40),

$$\frac{\left(\sum_{i:t_i \in b(t)} 1\right) - 1}{\sum_{h=1}^H \frac{V(A_{bh})}{V(A_{bh})+1}} = \frac{\frac{1}{H} \left(\left(\sum_{i:t_i \in b(t)} 1\right) - 1\right)}{\frac{1}{H} \sum_{h=1}^H \frac{V(A_{bh})}{V(A_{bh})+1}}$$

Using the strong law of large numbers, the numerator converges to

$$\Lambda^0(b(t)) \sum_A \alpha_{bA} \frac{V^0(A)}{V^0(A)+1}$$

almost surely. The denominator has already been shown to converge to $\sum_A \alpha_{bA} \frac{V(A)}{V(A)+1}$.

Thus, the result is proved. \square

Theorem 6. *For any given $t \in [0, L]$, let $\Lambda(t)$ denote the estimated cumulative arrival rate obtained by interpolating through $\Lambda(t_i)$ as defined in (31). If $\lim_{H \rightarrow \infty} \frac{|H_{bA}|}{H} = \alpha_{bA}$ for every A and b , and the set of collections of assortments in $\{S_b^+ : b = 1, \dots, |B|+1\}$ satisfies the identifiability condition in Theorem 4, then $\Lambda(t)$ converges to $\Lambda^0(t)$ almost surely for every t .*

Proof. According to our construction method of the cumulative arrival rate,

$$\sum_{i=1}^{I_t} \frac{l_i}{a_i} \leq \Lambda(t) \leq \sum_{i=1}^{I_t} \frac{l_i}{a_i} + \frac{l_{I_t+1}}{a_{I_t+1}}.$$

By Lemma 8, the extra term $\frac{l_{I_t+1}}{a_{I_t+1}}$ uniformly goes to zero almost surely. According to

Lemma 9, $\sum_{i=1}^{I_t} \frac{l_i}{a_i}$ converges to $\sum_b \frac{\Lambda^0(b(t)) \sum_A \alpha_{bA} \frac{V^0(A)}{V^0(A)+1}}{\sum_A \alpha_{bA} \frac{V(A)}{V(A)+1}}$ almost surely. Thus,

$$\Lambda(t) \xrightarrow{a.s.} \sum_b \frac{\Lambda^0(b(t)) \sum_A \alpha_{bA} \frac{V^0(A)}{V^0(A)+1}}{\sum_A \alpha_{bA} \frac{V(A)}{V(A)+1}}. \quad (41)$$

On the other hand, when the identifiability condition is satisfied, we know the estimate V converges to V^0 almost surely. Thus, $V(A)$ converges to $V^0(A)$ almost surely. Substitute $V(A_{hb})$ with $V^0(A_{hb})$ in (41), we have $\Lambda(t)$ converges to $\Lambda^0(t)$ almost surely for every t . \square

2.5 Numerical Examples

In this section, we illustrate the performance of the estimation algorithm described in the previous section through some numerical examples.

2.5.1 Effects of variations of assortments

As we have shown in Theorem 2, for the homogeneous arrival case, the necessary and sufficient condition for the identifiability of the arrival rate and the product attractivenesses is there should be variation in historical assortments, where the variation is in terms of the total attractiveness of an assortment. Actually, not only the existence of variation is critical for the identifiability, also the amount of the variation will greatly affect the estimation quality. The following experiment illustrates this idea.

In this set of experiments, two different assortments were offered: $S_1 = \{1\}$ and $S_2 = \{1, 2\}$. Each assortment will be offered for one unit of time. The attractiveness of product 1 is fixed at 1. The attractiveness of product 2 will be set to different values to reflect the variation between the two historical assortments. As the attractiveness of product 2 increases, the variation also increases. By varying the attractiveness of product 2, we can study the effects of the variation on the estimation quality.

Besides the variation among the historical assortments, the percentage of customers who purchase our products may also affect the estimation. We define the purchase rate α as Equation (42).

$$\alpha := \frac{\lambda \frac{V_1}{V_1+V_0} + \lambda \frac{V_1+V_2}{V_1+V_2+V_0}}{2\lambda} = \frac{\frac{V_1}{V_1+V_0} + \frac{V_1+V_2}{V_1+V_2+V_0}}{2} \quad (42)$$

To filter out the effects of purchase rate, the attractiveness of the no-purchase product is set so that the purchase rate is equal to some predetermined value.

In this set of experiments, we set $\alpha = 0.5$ and set $\lambda = 10,000$. V_2 will be increased from 0.1 to 2.0, at an increment of 0.1. We generate the random number of purchases for the corresponding product attractivenesses and arrival rate. Given the number of

purchases for each product and the assortment information, we can express the log-likelihood function in terms of V_1 , V_2 , V_0 , and λ as Equation (3). In the estimation, we set the attractiveness of product 1 as reference and scale V_1 equal to 1. We then maximize the log-likelihood function with respect to V_2 , V_0 , and λ . Under each parameter setting, 30 sets of synthetic data are generated. Correspondingly, we obtain 30 estimates. Figure 1a below shows the average percentage errors of the estimates of V_2 and λ under different values of V_2 . Figure 1b shows the variation of the estimation, which is defined to be the sample deviation of the estimation divided by the true value, under different values of V_2 . As we can see from the plots, the estimation of the arrival rate and V_0 is greatly affected by the value V_2 . When V_2 is small, which means the two historical assortments are very similar to each in terms of total attractiveness, the estimation of arrival rate λ and V_0 is inaccurate and very unstable. On the other hand, the estimation of V_2 is much less sensitive to the variation between the two assortments. In fact, the average percentage errors of V_2 under all values of V_2 are within 1%. Thus, airline companies should offer different assortments throughout the horizon not only as a way to respond to the market change, they also need the different assortments to improve their ability to learn the market.

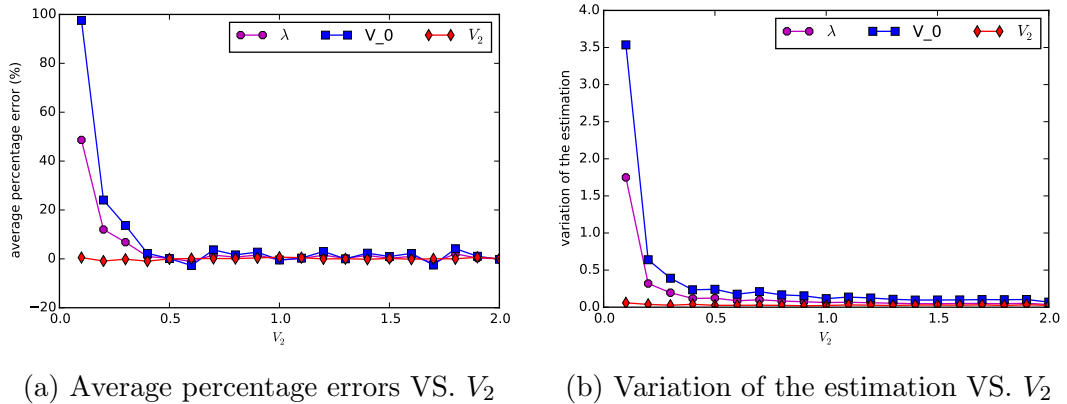
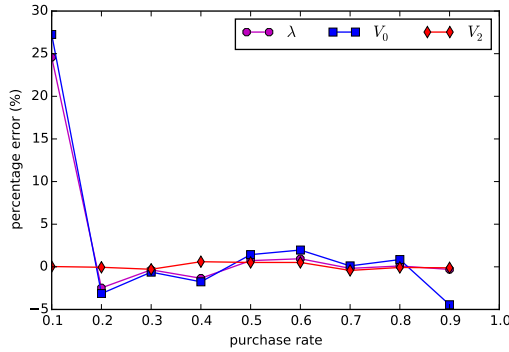


Figure 1: Quality of estimation VS. variation of assortments

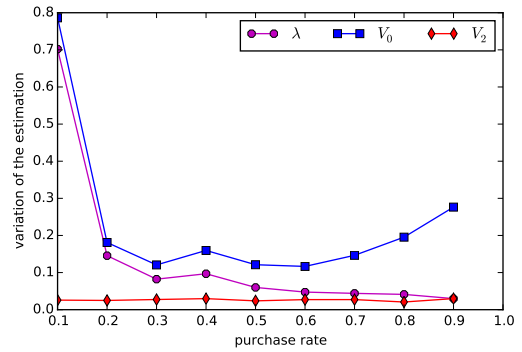
2.5.2 Effects of Purchase Rate

We know that the unobservation of the no-purchase option and the lack of variation among historical assortments lead to the unidentifiability of the arrival rate and the product attractiveness. When there is enough variation among historical assortments, it is also interesting to study the relationship between purchase rate as defined in (42) and the estimation quality. The purchase rate is very closely related to the market share of the company in practice.

In this experiment, two assortments were offered in history: $\{1\}$ and $\{1, 2\}$. Both Product 1 and Product 2 have attractiveness equal to 1. This guarantees there is enough variation between the two historical assortments. The attractiveness of the no-purchase option is varied so that the purchase rate could be set at different values. We also modify the arrival rate accordingly so that the expected number of purchases is equal to 5,000. Given V_1 , V_2 , V_0 , λ , and the assortment, we generate the random number of purchases of both products. We then use the purchase data and maximum likelihood method to estimate V_1 , V_2 , V_0 , and λ . Under each parameter setting, we repeat the data generation and estimation process for 30 times. The average percentage error and the variation of the estimate are as shown in Figure 2a and Figure 2b, respectively.



(a) Average percentage errors VS. purchase rate



(b) Variation of the estimation VS. purchase rate

Figure 2: Quality of estimation VS. purchase rate

As shown in the figures, when the purchase rate is below 10%, the estimates of V_0 and λ are inaccurate and unstable. The estimates of V_0 and λ quickly improves as the purchase rate increases. Thus, big airline companies have a natural advantage over the smaller ones in the sense that they can potentially learn the market conditions more accurately.

Since the information about the attractiveness of product 2 relative to product 1 is all captured in the sales data when assortment $\{1, 2\}$ is offered, the estimation quality of V_2 is not affected by the purchase rate.

2.5.3 Estimation of nonhomogeneous arrival rate

In this section, we deal with the case where the arrival rate function could be any integrable function and is unknown to the decision maker. We will use the estimation method developed in Section 2.4 to estimate the product attractiveness and the cumulative arrival rate function. In this set of experiments, we assume there are 4 products. Each product has attractiveness equal to 1. Note that if we combine two such products by summing up their attractiveness and summing up their purchase, we effectively create a new product with attractiveness equal to 2. Thus, assuming all products have the same attractiveness is without loss of generality. The selling horizon T is equal to 1 unit of time. The number of assortments during one selling horizon is a uniform random integer between 1 and 3. Given the number of assortments during the horizon, we generate a uniform random real number between 0 and 1 for each assortment. We then allocate the time length of the horizon to the assortments proportional to the random numbers. The availability of a product in an assortment is determined by a Bernoulli random variable with parameter 0.7. Note that during one selling horizon there may be some assortments that are the same with each other.

The nonhomogeneous arrival rate functions we test in the experiments include the

following:

- constant: constant arrival rate function, $\lambda(t) = \lambda$.
- pw2: piecewise constant rate function with 2 pieces,

$$\lambda(t) = \begin{cases} \frac{\lambda}{2}, & t \in [0, \frac{T}{2}] \\ \lambda, & t \in (\frac{T}{2}, T] \end{cases}$$

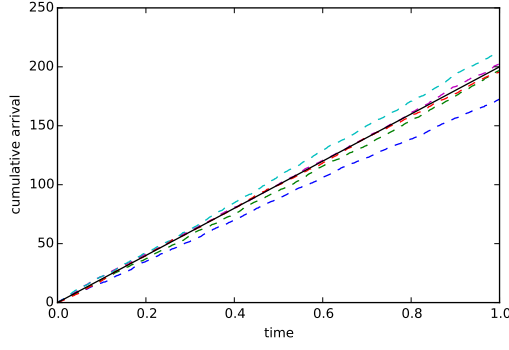
- linear: linear arrival rate function, $\lambda(t) = \lambda t$.
- cyclic: cyclic arrival rate function, $\lambda(t) = \lambda \frac{\sin(2\pi t)+1}{2}$.

Given the arrival rate function and the assortment throughout the horizon of all samples, we can define the expected purchase rate as a function of no-purchase attractiveness V_0 as following:

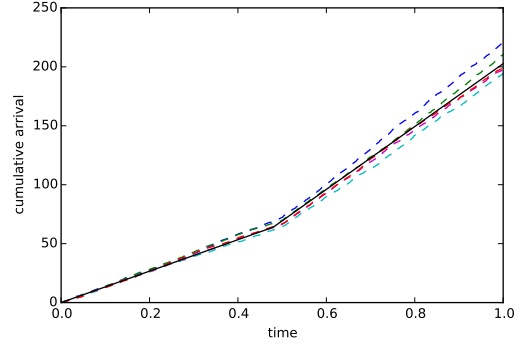
$$\alpha(V_0) = \frac{\sum_h \int_t \frac{V(A_h(t))}{V(A_h(t))+V_0} \lambda(t) dt}{\sum_h \int_t \lambda(t) dt}. \quad (43)$$

To filter out the effects of purchase rate, we set V_0 so that $\alpha(V_0)$ is equal to some predetermined value. After we determining V_0 , we can also set λ to control the expected total number of purchases from all samples.

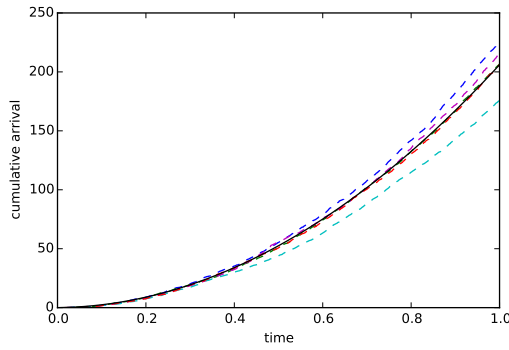
Figure 3a - Figure 3d plot the estimated cumulative arrival rate and the actual cumulative arrival rate when the purchase rate is 0.5 and the expected number of purchases from all samples is equal to 3,000. Given there are 30 samples, that is about 100 purchases from each sample, which is not a big number in practice. As we can see, the estimation method we developed in Section 2.4 can recover the actual arrival rate function reasonably well. Especially, the trend in the true arrival rate function can be correctly captured. For example, when the actual arrival rate is piecewise constant, as Figure 3b shows, in all 5 estimation experiments, our estimation results detect the kink point at $t = 0.5$, where the arrival rate changes.



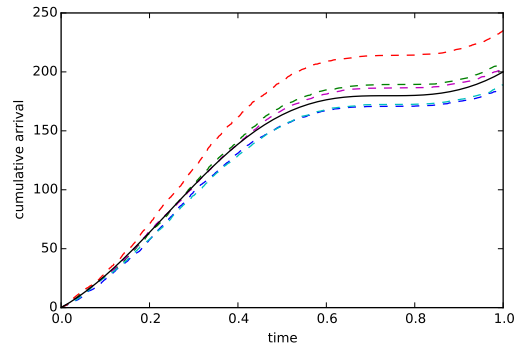
(a) The actual arrival rate is constant



(b) The actual arrival rate is piecewise constant



(c) The actual arrival rate is linear



(d) The actual arrival rate is cyclic

Figure 3: Estimation of cumulative arrival rate.

The solid line is the actual cumulative arrival rate and the 5 dashed lines are estimated from 5 different experiments.

In each experiment, we also evaluate the estimation error of the product attractiveness, the no-purchase attractiveness, and the arrival rate. To calculate the error of the attractivenesses, the estimated attractiveness of product 1 is normalized to be 1. All other attractivenesses are now relative to that of product 1. To evaluate the error of the estimate of the arrival rate function, we compare the estimated cumulative arrival rate with the true cumulative arrival rate at the end of the horizon. In most of our estimation experiments, if the cumulative arrival rate is over-estimated or under-estimated at the beginning of the horizon, it is over-estimated or under-estimated

throughout the horizon. Thus, the estimated cumulative arrival rate has its biggest error roughly at the end of the horizon. We calculate the average percentage error of V_2 , V_3 , V_4 , V_0 , and $\Lambda(T)$. We also calculate the sample standard deviation of the percentage errors of these quantities. The results are as shown in Table 1. The small average percentage errors in the table indicate that the estimation method we developed has small bias. The sample standard deviations of percentage errors of V_2 and V_3 are much smaller than that of V_0 and $\Lambda(T)$. This is because unlike V_0 and $\Lambda(T)$, the information of the relative popularity among product 1, product 2, and product 3 is not lost due the unobserved no-purchase data.

2.6 Conclusion

In this chapter, we study the estimation of the choice model parameters and the market level arrival rate using only the sales data. We give sufficient and necessary conditions under which the estimation parameters are identifiable. Our conditions cover the cases when the arrival rate is constant, when the arrival rate is piecewise constant, and when the arrival rate is any integrable rate function. These conditions are established for the first time in literature and they can help to resolve the confusion regarding the estimation problem with censored data. The confusion has been existing for a long time both in industry and research society. We also develop a novel estimation algorithm that can recover any arrival rate function when the identification conditions are satisfied. The algorithm not only has theoretical convergence guarantee, it also has good performance with limited data. The algorithm can be very useful for practitioners to discover the arrival pattern in the market.

Table 1: Quality of estimation for nonhomogeneous arrival rates

The average percentage error (the first element of Column 2 - Column 6) and the sample standard deviation (the second element of Column 2 - Column 6) of the percentage errors for V_2 , V_3 , V_4 , V_0 , and $\Lambda(T)$. The first element of the case column denotes the purchase rate and the second element denotes the true arrival rate function. The expected total number of purchases from all the samples is 3,000.

case	V_2	V_3	V_4	V_0	$\Lambda(T)$
(0.3, constant)	(-0.68, 5.61)	(-1.49, 5.61)	(-1.37, 4.91)	(-19.66, 21.91)	(-13.25, 14.15)
(0.5, constant)	(4.01, 2.89)	(0.59, 2.89)	(5.87, 6.42)	(-0.93, 12.46)	(-2.00, 6.60)
(0.7, constant)	(-0.23, 7.96)	(0.69, 7.96)	(-1.72, 5.00)	(-4.14, 25.82)	(-1.25, 7.53)
(0.3, pw2)	(5.29, 2.93)	(2.74, 2.93)	(3.60, 3.95)	(1.79, 19.96)	(-2.36, 14.45)
(0.5, pw2)	(1.43, 6.15)	(0.93, 6.15)	(-2.53, 5.02)	(3.71, 15.28)	(0.94, 4.74)
(0.7, pw2)	(1.57, 2.99)	(4.54, 2.99)	(3.16, 3.01)	(-5.75, 9.18)	(-3.15, 2.55)
(0.3, linear)	(1.98, 3.85)	(3.32, 3.85)	(1.74, 3.49)	(-15.54, 19.07)	(-12.81, 10.96)
(0.5, linear)	(-0.10, 4.56)	(2.35, 4.56)	(4.47, 4.28)	(8.61, 15.83)	(0.07, 8.19)
(0.7, linear)	(1.38, 8.38)	(2.87, 8.38)	(5.24, 9.33)	(-0.51, 16.62)	(-3.32, 6.76)
(0.3, cyclic)	(-3.89, 1.97)	(-1.66, 1.97)	(-2.20, 3.06)	(12.53, 19.26)	(10.88, 12.25)
(0.5, cyclic)	(-0.89, 5.46)	(-3.45, 5.46)	(-0.41, 3.22)	(0.90, 15.36)	(2.12, 8.77)
(0.7, cyclic)	(1.76, 7.11)	(-1.64, 7.11)	(0.11, 4.94)	(-6.32, 29.90)	(-1.74, 7.86)

CHAPTER III

REVENUE MANAGEMENT UNDER CHOICE MODEL WITH BUY-DOWN EFFECTS

3.1 Introduction

Revenue management techniques have been used extensively in many industries, e.g., the airline industry, the hospitality industry, the car rental industry and so on. When customers want to purchase some product in these industries, they usually first inquire what products are available. Then they evaluate all the available alternatives and make the choice among them. Such a decision process is well suited for a discrete choice model. Also, most of the discrete choice models can incorporate customers' sensitivities to various attributes of the products. Because of these advantages, using discrete choice models to describe customers' purchasing behavior has been one clear trend recently in both the revenue management literature and the revenue management industry.

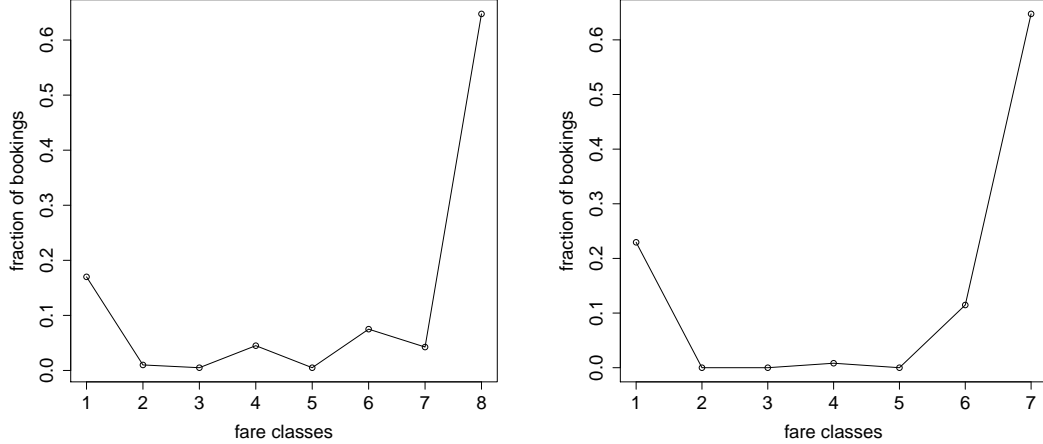
One immediate question that revenue management researchers and practitioners need to answer is how to determine the product availability given the customer choice behavior so that the revenue performance can be maximized. The determination of the availability is often referred to as the assortment optimization problem or the dynamic optimization problem, depending whether capacity constraint and the temporal dimension are introduced. Various studies have been devoted to address the assortment optimization problem under different customer choice models (see Section 3.2 for more details).

In this study, we will incorporate the buy-down effects into the customer choice

model and develop algorithms for how to solve the corresponding assortment optimization problem both statically and dynamically. The buy-down effects refer to the phenomenon that the fact an alternative being the cheapest one among the choice set will dramatically increase its attractiveness to the customer. Such effects are intuitive to understand. When several products have more or less the same attributes except the price, most of the purchases, if there are any, will go to the one with the lowest price. Even when the products have different attributes, being the cheapest will often send the product to the top of the search results in many search engines. The top position usually attracts the most eye impressions and thus will result in more purchases.

As an example, Figure 4a and 4b gives the distribution of the sales of all the available fare classes on a same flight when fare class 8 and 7 are the cheapest available, respectively. The fare classes in the plot are sorted in descending order of prices and fare class 1 is the most expensive. The data of the plot are from a real world company for whom we did a revenue management project. As we can see from the plot, when fare class 8 was the cheapest, most of the bookings occurred to fare class 8. When fare class 8 was taken down and fare class 7 became the cheapest available, most of the bookings occurred to fare class 7. Being the cheapest greatly boosted the sales of fare class 7. The relative sales between fare class 6 and fare class 7 before and after fare class 7 became the cheapest available are also very different. In the later case, fare class 7 became much more popular relative to fare class 6. This is in sharp contrast with the MNL choice model, which will assert the relative popularity between fare class 6 and 7 be the same as long as both products are in the assortment.

When the buy-down effects are extremely strong, customers ignore the difference of any other attributes and only purchase the cheapest available product. Such a phenomenon will be called 100% buy-down effects. For example, in airline industry, many low-cost airlines adopt a simple product structure so the price is essentially the



(a) Sales distribution when fare class 8 is the cheapest available (b) Sales distribution when fare class 7 is the cheapest available

Figure 4: Sales distribution when different fare classes are the cheapest available

only difference among different seats on the same flight. In this case, customers may still trade off among different flights which have different departure times, but only the cheapest available seats in each flight will be considered. In this case, the airline company only needs to post one price tag for each flight, which is in fact the current practice for most of the low-cost airline companies.

We remark that although we motivate the customer choice model with 100% buy-down effects from the revenue management applications with undifferentiated fares, the problem arises naturally in retailing as well. In retail industry, it does not make sense to post multiple prices for one product. This can be viewed as if the retailer post multiple prices for one product, all the customers will only stick to the cheapest price if they buy. Given that customers still choose from similar products, the assortment optimization problem under choice model with 100% buy-down effects can also be viewed as a special case of the pricing problem, where there is only a finite menu of prices to choose from.

Similar with the general buy-down effects, the 100% buy-down effects cannot be captured by a traditional choice model like MNL or NL. These choice models always

assign a positive choice probability for each product in the choice set. Modifying the existing choice model so it can be used with 100% buy-down effects will be useful in practice and solving the resulting dynamic assortment optimization problem have been a challenge for airline industry and hospitality industry for quite a long time.

One important feature of the buy-down effects is that the attractiveness of a product is not only a function of its own attributes, it is also dependent on the attributes of other products in the same choice set. Such dependency cannot be captured by any of the traditional choice models like multinomial logit model (MNL) or nested logit model (NL). Ignoring the buy-down effects in demand estimation would result in lower predictions of more expensive products and higher predictions of cheaper products. Such forecast further exacerbates the problem by encouraging less capacity be reserved for the more expensive products. The resulting behavior is known as the “spiral-down” effect [11]. One way to modify the MNL choice model so that it can reflect the buy-down effects is through the introduction of a special parameters for the cheapest product (see [13] for details). With such modification, the estimation procedure of the MNL model is the same as the estimation of a general MNL model. However, both the static and the dynamic assortment optimization problem with such modified MNL choice model will be more complicated. Developing an efficient algorithm will be one of the main tasks of this chapter.

Another important problem that will be tackled in this section is related to competition. In revenue management with customer choice model, there is usually a special alternative in the choice set which is often referred to as the no-purchase alternative. The no-purchase alternative is an aggregation of all the alternatives that are not offered by the company itself. It includes not only the alternative that a customer really does not purchase anything and leaves the system, but also all the alternatives from competitors. In reality, it is difficult to know what alternatives the competitors

will offer in advance. Usually, the best a company can do is to have some distribution estimation about competitors' offerings. This makes the no-purchase alternative random. How to solve the assortment optimization problem with buy-down effects and random no-purchase alternative will be another focus of the section.

3.1.1 Contributions

The main contributions of this chapter include the following.

- We develop a compact sales based linear programming model (SBLP) for the dynamic assortment optimization problem under MNL choice model with general or 100% buy-down effects. In the SBLP models, the decision variables correspond to the sales of each product. The number of decision variables in the SBLP with 100% buy-down effects is quadratic in the number of products.
- We develop efficient polynomial algorithms that convert the sales solution of the SBLP model to the corresponding assortment solution. The assortment solution is more suitable for implementation.

Neither our definition of the decision variables in the SBLP formulation nor the conversion algorithm from sales solution to assortment solution is a straightforward extension of those in [23]. In fact, when there is buy-down effects, we have to define multiple sales variables for one product, instead of just one sales variable for each product in [23].

- We develop polynomial time algorithms to solve the static assortment optimization problem with 100% buy-down effects and random no-purchase alternative. The assortment optimization algorithm can then be used to solve the dynamic assortment optimization problem through column generation algorithm.
- When the attractiveness of the no-purchase alternative in the MNL model is uncertain, and the customers have general buy-down behavior, we prove that under

very mild conditions, the optimal assortment has some nice nesting structure. The nesting structure allows us to reduce the assortment optimization problem with *general buy-down effects* and random no-purchase alternative to the assortment optimization problem with *100% buy-down effects* and random no-purchase alternative. Thus, both the static and dynamic assortment optimization problem with general buy-down effects and random no-purchase alternative can be solved efficiently.

3.2 *Related literature*

The 100% buy-down effects have been noticed for a long time in practice. In airline industry, the customers who only purchase the cheapest available product are termed as priceable demand and the customers who care about other attributes of the products and may choose a more expensive product even though a cheaper one is available are termed yieldable demand. [9] first noticed that as the competition among airline companies increase and the prevalence of the Internet, more and more demands will be of the priceable type and [9] suggest that the revenue management models should be modified to accommodate this change.

As we have said in the introduction, the assortment optimization problem with 100% buy-down effects can be viewed as a pricing problem. Most existing literature implicitly discuss the 100% buy-down effects through the pricing model. [25] consider a multiproduct dynamic pricing problem where the demand for each product is a stochastic point process. The intensity of the point process for one product is a function of the vector of prices for all the products. The decision variables are continuous. If customers choose the products at different prices according to a choice model, the intensity functions can be defined accordingly. Two asymptotically optimal heuristics based on a deterministic model are developed in [25]. [60] study a dynamic problem for substitutable flights. The probability that a customer chooses

one particular flight is a general function of the price vector. Various bounds of the dynamic programming problem and some heuristic methods are developed.

To our best knowledge, the revenue management problem under choice model with general buy-down effects has not been discussed in existing literature. But the problem is of great importance in practice. As [9] argued, the real demand is a mix of yieldable demand and priceable demand. On the other hand, the assortment and pricing optimization problem under customer choice model without the buy-down effects has been studied intensively.

The static assortment optimization problem under choice model can be an important building block to develop efficient algorithms for the dynamic assortment optimization problem. [14] study the assortment optimization problem under MNL model with totally unimodular constraints. The study is of particular interest to us since the 100% buy-down effects can be enforced by allowing only one price for one product. This can be expressed as totally unimodular constraints. [16] study the pricing problem under NL model, of which MNL is a special case, with quality consistency constraints. When the prices are allowed to be continuous, [34] study the multiproduct pricing problem under MNL and NL choice model. [26] study the multiproduct pricing problem under NL model with product-differentiated price sensitivities. [43] study the multiproduct pricing problem under NL model with price bounds. [45] develop an efficient algorithm for the assortment optimization under the MNL model with capacity constraint.

[44] discuss the assortment optimization problem under the MNL model with random choice parameters. The problem is more general than one of our problems in that the attractivenesses of both company's own products and the no-purchase alternative are random. It is shown in [44] that although the assortment optimization problem is easy to solve under certain conditions, the problem is generally NP-hard. To solve the same problem, [37] give a branch-and-cut algorithm and [10] propose

a greedy heuristic algorithm. [46] consider a robust version of assortment optimization problem when the parameters are in a compact uncertainty set and show that an efficient algorithm can be developed. When the consideration sets are nested to each other in the MNL model, [20] prove that the corresponding assortment optimization problem is NP-complete. A fully polynomial time approximation algorithm is then proposed. [15] give the necessary and sufficient conditions under which the assortment optimization with nested logit model is polynomially solvable. When the conditions are satisfied, [33] give an efficient algorithm to solve the assortment optimization problem with d -level nested logit model. [17] developed fully polynomial time approximation schemes (FPTAS) for assortment optimization problem with the nested logit model and the mixed MNL model. [18] consider the assortment optimization problem under Markov choice model with capacity constraint. [57] consider the assortment optimization problem under general attraction model with capacity constraint.

The dynamic assortment (pricing) optimization problem studies how to determine the assortment (prices) throughout the planning horizon so that the consumed resources are within capacity limit, given that both the customer arrival process and the customer choice behavior are random. Although the problem can be modeled as a Markov decision problem, the resulting model is usually too large to solve. As an alternative, [22] propose a deterministic approximation model where the stochastic quantities are replaced with their expected values and the capacity and the demand are treated as continuous. The resulting linear programming model is called choice based deterministic linear programming (CDLP). In CDLP, the number of decision variables is equal to the number of potential offer sets, which is exponential in the number of products. Thus, a column generation algorithm is often needed. The subproblem of the column generation algorithm is an assortment optimization problem. According to the equivalence between separation and optimization in [27], if one can

show that the assortment optimization problem with certain choice model can be solved efficiently, the CDLP with the same choice model can also be solved efficiently. Thus, many of the static assortment optimization researches we just discussed can also be umbrellaed under this branch.

One particularly interesting research about the dynamic assortment optimization problem we found is [23], which develop a sales based linear programming (SBLP) model to approximate the dynamic assortment optimization problem. In the SBLP model, the decision variables are the sales quantities of the products, and the number of decision variables is only linear in the number of products. This is a significant improvement over the CDLP model. [23] also show that the CDLP and SBLP solutions can be converted from each other within polynomial time. Thus, the two formulations are equivalent. Developing SBLP formulations for other choice models has received a lot of attention recently. [21] develop an SBLP model under the Markov chain choice model. A polynomial time algorithm is also developed to convert the sales solution to the assortment solution of CDLP.

3.3 Dynamic Assortment Optimization with General Buy-down Effects

In this section, we describe the choice model we use that incorporates the buy-down effects and propose a compact sales based model for the dynamic assortment optimization problem. An efficient algorithm is developed to convert the sales solution to an implementable solution.

3.3.1 MNL model with buy-down effects

We now describe a modified MNL choice model that can handle the buy-down effects. Let $\mathcal{J} = \{1, \dots, J\}$ be the set of products that the decision maker can offer. The collection of sets J_1, \dots, J_M form a partition of \mathcal{J} . Let $J(j)$ denote the subset that product j belongs to. In airline industry, the partition could be according to

itineraries and J_m contains all the fare classes on the itinerary m . In retailing, the partition could be according to products and J_m contains all the possible prices tags that can be applied to product m .

Within each subset, if the assortment is not empty, the cheapest available product will receive the buy-down effects. When product j is the cheapest available among $J(j)$, its attractiveness to the customer is w_j . Otherwise, the attractiveness is v_j . We assume $w_j \geq v_j$. Let $I(j, A) = 1$ if product j is the cheapest available among $J(j)$ when the assortment is A . Otherwise, $I(j, A) = 0$. Denote the attractiveness of the no-purchase option as v_0 . When the assortment is A , a customer chooses product $j \in A$ with probability equal to:

$$P_{j:A} = \frac{w_j I(j, A) + v_j (1 - I(j, A))}{v_0 + \sum_{j' \in A} [(w_{j'} I(j', A) + v_{j'} (1 - I(j', A)))]}. \quad (44)$$

If $j \notin A$, then $P_{j:A} = 0$.

3.3.2 The decision problem and its MDP formulation

Product j 's revenue is r_j . Without loss of generality, we assume that products in the same subset are ordered in descending order of their revenues. There are F resources the company can use to assemble the products. As an example, the resources refer to the flights legs in airline industry. Product j will use B_{jf} units of resource f . Let B_j denote the vector $(B_{jf})_{f=1, \dots, F}$. The capacity of resource f is b_f . The selling horizon is divided into T discrete-time periods. Period T denotes the end of the horizon. In each period, there is one customer arrival with probability λ and no customer arrival with probability $1 - \lambda$. Note that the arrival rate λ is allowed to be time-dependent and all the following analysis still holds true. We ignore the time dependency for the ease of exposition. Customers choose the product from the assortment according to Equation (44).

The decision maker's problem is to dynamically determine the assortment to offer in each period so the expected revenue can be maximized. The problem can be

modeled as a Markov decision process. Let $s = (s_1, \dots, s_F)$ be a vector whose f^{th} element denotes the remaining capacity of resource f . Let $v_t(s)$ denote the maximum expected revenue-to-go given that the state in period t is s . The Bellman equations for the MDP are as following:

$$v_t(s) = \max_A \left\{ \lambda \sum_{j \in A} P_{j:A} (r_j + v_{t+1}(s - B_j)) + (1 - \lambda + \lambda P_{0:A}) V_{t+1}(s) \right\}, \forall t, s. \quad (45)$$

The boundary conditions are $V_T(s) = 0$ for all s .

Although the backward induction algorithm can be used to solve the MDP problem (45), the large state space and decision space make the algorithm computationally intractable. For example, suppose an airline network has 10 flights and 20 itineraries, and each flight has 100 seats. Such a network is a quite small in practice. In the dynamic assortment optimization problem for this network, there will be 100^{10} states and 2^{20} possible decisions. On the other hand, even problem (45) can be solved exactly, the resulting policy is a big lookup table. Such a policy is difficult to store and implement in practice. To overcome these difficulties, some approximation methods need to be developed.

3.3.3 CDLP formulation

One popular deterministic approximation to the MDP formulation (45) is to replace all the stochastic quantities with their expected values and to treat the capacity and demand as continuous (see [22, 36]). The resulting model is called choice based deterministic linear programming formulation (CDLP).

Let the decision variable $u(A)$ denote the fraction of time when assortment A is offered. Let $R(A)$ and $Q_f(A)$ denote the expected revenue and the expected consumption of resource f when assortment A is offered throughout the horizon. Using

the definition of choice probability as in (44), their definitions are as following:

$$R(A) = \lambda T \sum_{j \in A} P_{j:A} r_j \quad (46)$$

$$Q_f(A) = \lambda T \sum_{j \in A} P_{j:A} B_{jf} \quad (47)$$

With this notation, the CDLP formulation is as following:

$$\begin{aligned} \max_u \quad & \sum_A R(A) u(A) \\ \text{s.t.} \quad & \sum_A Q_f(A) u(A) \leq b_f \quad \forall f \\ & \sum_A u(A) \leq 1 \\ & u(A) \geq 0 \quad \forall A \end{aligned} \quad (\text{CDLP})$$

The first set of constraints in (CDLP) guarantee that the resource consumption should not exceed their capacity limit. The unit constraint says that we can only operate within the planning horizon.

The CDLP model offers an upper bound to the original MDP problem (45). [36] proves that as the arrival and capacity scale to infinity proportionally, the upper bound is asymptotically tight. In fact, the CDLP formulation is generic and it can be used to accommodate any choice model. One only needs to replace the definition of choice probability $P_{j:A}$ in (46) and (47). One potential drawback of the CDLP formulation is that the number of decision variables in the CDLP is exponential in the number of products. Often, a column generation algorithm is used to solve the CDLP formulation.

The gist of the column generation algorithm is to figure out the promising columns given the current dual variables. Suppose the dual variable corresponding to the capacity constraint f is π_f and the dual variable corresponding to the unit constraint is μ . The subproblem of the column generation algorithm is as following:

$$\max_A R(A) - \sum_f \pi_f \cdot Q_f(A) - \mu = \max_A \lambda T \sum_{j \in A} P_{j:A} \left(r_j - \sum_f B_{jf} \pi_f \right) - \mu. \quad (48)$$

The essence of problem (48) is an assortment optimization problem with modified revenue of product j equal to $r_j - \sum_f B_{jf} \pi_f$. If we can solve this assortment optimization problem within polynomial time, according to [27], the CDLP can also be solved within polynomial time. Although the column generation algorithm can be used to solve the dynamic assortment optimization problem, we will use the SBLP formulation, which is much more compact and easier to solve than the CDLP formulation. We will come back to the column generation algorithm when the no-purchase alternative is random.

3.3.4 SBLP formulation

We now propose an SBLP formulation for the revenue management problem under customer choice model with buy-down effects. Let x_j denote the sales of product j when it is the cheapest available among $J(j)$. Let $x_{j'}^j$ denote the sales of product j' when product j is the cheapest available among $J(j')$. Since the revenues of the products are of decreasing order in each subset, for each j , we only need to define $x_{j'}^j$ for $j' < j$ and $j' \in J(j)$. Let x_0 denote the number of customers who ‘purchase’ the no-purchase alternative. The SBLP formulation is as following:

$$\max \sum_{j \in \mathcal{J}} \left(r_j x_j + \sum_{j' \in J(j), j' > j} r_{j'} x_{j'}^j \right) \quad (49a)$$

$$\text{s.t.} \sum_{j \in \mathcal{J}} \left(x_j + \sum_{j' \in J(j), j' > j} x_{j'}^j \right) + x_0 = \lambda T \quad (49b)$$

$$\sum_{j \in \mathcal{J}} \left(B_{jf} x_j + \sum_{j' \in J(j), j' > j} B_{j'f} x_{j'}^j \right) \leq b_f \quad \forall f \in F \quad (49c)$$

$$\frac{x_{j'}^j}{v_{j'}} \leq \frac{x_j}{w_j} \quad \forall j \in \mathcal{J}, \forall j' \in J(j) \text{ and } j' < j \quad (49d)$$

$$\sum_{j \in J_m} \frac{x_j}{w_j} \leq \frac{x_0}{v_0} \quad \forall m \in \{1, \dots, M\} \quad (49e)$$

$$x \geq 0$$

The flow balance constraint (49b) guarantees that each arrival customer either buy some product or leave the system without purchasing anything. Constraint (49c) guarantees that the capacity of each resource will not be exceeded. The quantity $\frac{x_{j'}^j}{v_{j'}}$ can be viewed as a normalized approximation of the time that product j is offered when some other product j' in $J(j)$ is the cheapest available. Similarly, $\frac{x_j}{w_j}$ can be viewed as an approximation of the time that product j is offered as the cheapest available product in $J(j)$. If whenever product j is offered and it is the cheapest available product in $J(j)$, product j' ($j' < j$ and $j' \in J(j)$) is also offered, then constraint (49d) will hold with equality. The less than or equal to relation in constraint (49d) allows product j' not be offered all the time when j is the lowest available. Since the no-purchase option is always available, constraint (49e) makes sure that at any time, either the assortment in J_m is not empty and some product is the cheapest in it or no alternative from J_m is offered.

All constraints in the formulation (49) are intuitively necessary. But it is still unclear whether these constraints are sufficient to guarantee that a sales solution to (49) is actually achievable by offering certain assortments for certain amount of

time, e.g. a CDLP solution. On another hand, the sales quantities are difficult to implement in practice. For implementation purpose, we also need to convert an SBLP solution to a CDLP solution. We address these two questions by offering an conversion algorithm as Algorithm 1.

Algorithm 1 Algorithm of converting an SBLP solution to a CDLP solution

Input:

An SBLP solution \mathbf{x}
Set of products \mathcal{J} and its partition J_1, J_2, \dots, J_M
Attractiveness of product j : v_j and w_j , for all $j \in \mathcal{J}$
Attractiveness of the no-purchase alternative: v_0 .
Total arrival: $\lambda \cdot T$

Initialize:

Set $u(A) = 0$ for all $A \subset \mathcal{J}$ and $A \neq \emptyset$
Set $x = \mathbf{x}$
Set $i = 0$

- 1: **if** $x_j = 0$ for all $j \in \mathcal{J}$ **then**
- 2: Set $u(\emptyset) = 1 - \sum_{i'=1}^i u(A_{i'})$. Output u as the CDLP solution.
- 3: **for** $m = 1, \dots, M$ **do**
- 4: **if** $x_j = 0$ for all $j \in J_m$ **then**
- 5: Define $Y_m = -\infty$
- 6: **else**
- 7: Pick any $j(m) \in J_m$ such that $x_{j(m)} > 0$. Define

$$Y_m = \min \left\{ \left\{ \frac{x_{j(m)}}{w_{j(m)}} \right\} \cup \left\{ \frac{x_j^{j(m)}}{v_j} : j' < j(m), j' \in J_m, x_{j'}^{j(m)} > 0 \right\} \right\}$$

- 8: Compute $I = \min \{Y_m : Y_m > 0\}$
- 9: Define

$$A_{i+1} = \bigcup_{m: Y_m > 0} \left(\{j(m)\} \cup \{j : x_j^{j(m)} > 0\} \right)$$

- 10: Define

$$u(A_{i+1}) = \frac{I \left(\sum_{m: Y_m > 0} \left(w_{j(m)} + \sum_{j: x_j^{j(m)} > 0} v_j \right) + v_0 \right)}{\lambda T}$$

- 11: **for** $m = 1, \dots, M$ **do**
- 12: **if** $Y_m > 0$ **then**
- 13: Update $x_{j(m)}$ as

$$x_{j(m)} = x_{j(m)} - I \cdot w_{j(m)}$$

- 14: **for** all j with $j < j(m)$ and $x_j^{j(m)} > 0$ **do**
- 15: Update $x_j^{j(m)}$ as

$$x_j^{j(m)} = x_j^{j(m)} - I \cdot v_j$$

- 16: Update $i = i + 1$. Go back to step 1
-

Figure 5 provides a graphical illustration of Algorithm 1. In the example of Figure 5, $M = 2$, $J_1 = \{1, 2, 3\}$, and $J_2 = \{4, 5\}$. In Figure 5, interval $\frac{x_j}{w_j}$'s within the same J_m are placed adjacently. For $j' < j$ and $j' \in J(j)$, interval $\frac{x_{j'}}{v_{j'}}$ is placed underneath interval $\frac{x_j}{w_j}$ and the two intervals are aligned to left. Because of Constraint 49d, interval $\frac{x_{j'}}{v_{j'}}$ will be completely underneath $\frac{x_j}{w_j}$. For each $\frac{x_j}{w_j}$, we can find its overlap with all $\frac{x_{j'}}{v_{j'}}$'s beneath it. If there are no intervals underneath $\frac{x_j}{w_j}$, the overlap is defined to be $\frac{x_j}{w_j}$ itself. This overlap is actually the Y_m in step 7 of Algorithm 1. We further take the minimum of Y_m 's and we obtain interval I_1 in Figure 5. The assortment corresponding to I_1 is $A_1 = \{1, 4\}$ with $u(A_1) = \frac{(w_1+w_4+v_0)I_1}{\lambda T}$. After A_1 is generated, all subintervals under I_1 are removed and a new iteration begins. The second assortment is $A_2 = \{1, 4, 5\}$ with $u(A_2) = \frac{(w_1+w_5+v_4+v_0)I_2}{\lambda T}$. Following this logic, 9 different nonempty assortments will be generated. Note that the assortment corresponding to interval I_{10} in Figure 5 is empty.

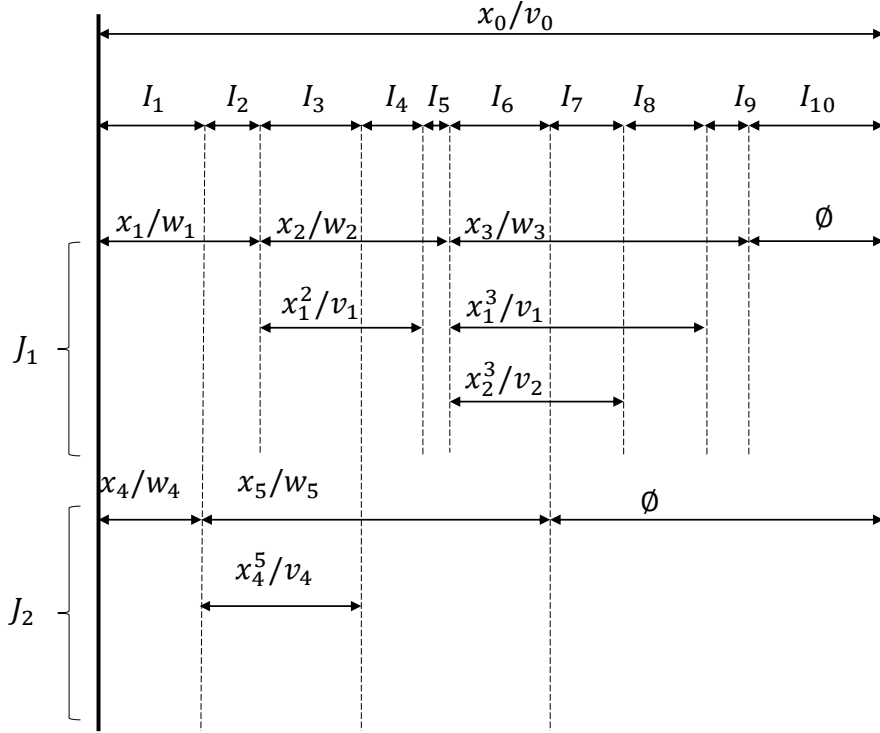


Figure 5: Conversion of a feasible SBLP solution to a feasible CDLP solution.

The following theorem says that Algorithm 1 will terminate after finite steps.

Theorem 7. *Given a feasible SBLP solution x , Algorithm 1 will terminate after at most N iterations, where N is the number of positive elements in x .*

Proof of Theorem 7. In each iteration, because of the definition of I , at least one of the positive elements of x will be reduced to 0. Thus, after at most N iterations, x will be reduced to 0 and the algorithm stops. \square

Below are some other properties of Algorithm 1.

Lemma 10. *Given a feasible SBLP solution, the following results hold for each iteration i :*

1. x satisfies constraints (49d) and (49e);
2. The expected sales resulted from offering assortment A_{i+1} for $u(A_{i+1}) \cdot T$ unit of time are equal to the sales subtracted from x in iteration i .

Proof of Lemma 10. Suppose that at the beginning of an iteration, x satisfies constraints (49d) and (49e). Thus,

$$\begin{aligned} \frac{x_{j'}^j}{v_{j'}} &\leq \frac{x_j}{w_j} \quad \forall j \in \mathcal{J}, \forall j' \in J(j) \text{ and } j' < j \\ \sum_{j \in J_m} \frac{x_j}{w_j} &\leq \frac{x_0}{v_0} \quad \forall m \in \{1, \dots, M\} \end{aligned}$$

During iteration i , one of the $\frac{x_j}{w_j}$'s will be subtracted by I , and the same with $\frac{x_0}{v_0}$. Thus, constraint (49e) still holds in the next iteration.

If I has intersection with $\frac{x_j}{w_j}$, then both $\frac{x_{j'}^j}{v_{j'}}$ and $\frac{x_j}{w_j}$ will be subtracted by the same quantity I , leaving constraint (49d) valid in the next iteration. If I has no intersection with $\frac{x_j}{w_j}$, constraint (49d) will be the same in the next iteration and still holds trivially.

Given the above inductive results and the condition that the starting solution \mathbf{x} is feasible (thus satisfies constraints (49d) and (49e)), the first part of the lemma is established.

The second part of the result is straightforward if we substitute the definition of $u(A)$ to calculate the expected sales. \square

Now we are ready to show the equivalence between the SBLP formulation and the CDLP formulation.

Theorem 8. *Given any feasible solution x for (49), a feasible solution u for (CDLP) can be computed in polynomial time, and vice versa, such that the two solutions represent the same number of bookings for each product and have the same objective values.*

Proof of Theorem 8. First we show that for any feasible solution u for (CDLP), there is a feasible solution x for (49) representing the same number of bookings and having the same objective value. Given u , let x denote the resulting number of bookings. For example, if $j \in J_m$, then $x_j = \sum_{\{A: I(j,A)=1\}} \lambda TP_{j:A} u(A)$. Then, x satisfies (49b). Since u is feasible for (CDLP), it follows that x satisfies (49c). Since bookings x_j^j take place only when j is the cheapest available product in the subset $J(j')$, (49d) follows from the IIA property of the MNL choice model. Finally, since the alternative not to book anything is always available, (49e) is satisfied by x .

For the other direction, given x , we use Algorithm 1 to convert x to a CDLP solution. According to Lemma 10, the corresponding CDLP solution generates the same sales as x . \square

The SBLP formulation has $\frac{J(J+1)}{2}$ decision variables and $1 + F + \frac{J(J+1)}{2} + M$ constraints, both of which are quadratic in the number of products and linear in the number of resources. Thus, the SBLP formulation is much more compact than the CDLP formulation. Given an SBLP solution, Algorithm 1 provides an efficient conversion algorithm to a CDLP solution.

3.4 *Dynamic Assortment Optimization with 100% Buy-down Effects*

In this section, we discuss the dynamic assortment optimization when the buy-down effects are extreme and customers always purchase the cheapest available product

within a subset, if they ever choose to purchase anything. 100% is a prominent phenomenon in airline industry. For many leisure airline markets, customers always only choose the lowest available tickets on a flight. In hotel industry, a hotel needs to decide one price tag for each room type, this can be interpreted as if the prices above the tag are still available, but customers always choose the lowest available price. Of course, in the above two examples, customers may still choose among different flights or different room types. Thus a customer choice model with 100% buy-down effects is appropriate for the application and has attracted lots of attention in practice.

Although 100% buy-down effects can be viewed as a special case of the general buy-down effects, due to its practical importance, we separately give the model and the solution methods.

As in Section 3.3.1, let $\mathcal{J} = \{1, \dots, J\}$ denote the set of products which the company has. The collection of sets J_1, \dots, J_M form a partition of \mathcal{J} . Customers only consider the cheapest available product from each subset J_m . Subset J_m can be all fare classes from the same flight in airline industry, or all price tags for the same room type in hotel industry. The attractiveness of product j is w_j . Let $J(j)$ denote the subset that product j belongs to. Let $I(j, A) = 1$ if product j is the cheapest available in $J(j)$ when assortment A is offered. Let $I(j, A) = 0$ otherwise. The choice probability that a customer chooses product j when assortment A is offered is as following:

$$P_{j:A} = \frac{w_j I(j, A)}{v_0 + \sum_{j' \in A} w_{j'} I(j', A)}. \quad (50)$$

There are F resources the company can use to assemble the products. Product j will use B_{jf} units of resource f . Let B_j denote the vector $(B_{jf})_{f=1, \dots, F}$. The capacity of resource f is b_f . The selling horizon is divided into T discrete-time periods. Period T denotes the end of the horizon. In each period, there is one customer arrival with probability λ and no customer arrival with probability $1 - \lambda$.

With such setup, we can develop an MDP model for the dynamic assortment

optimization problem with 100% buy-down effects. The only difference is that the choice probability in Equation (45) now refers to (50). Similarly, we can define $R(A)$ and $Q_f(A)$ using the new choice probability. With the updated $R(A)$ and $Q_f(A)$, the formulation CDLP can also be used as a deterministic approximation to the MDP.

For the CDLP formulation with 100% buy-down effects, an efficient column generation algorithm can be developed.

Suppose the dual variable corresponding to the the first constraint (the unit constraint) is π . The dual variable corresponding to recourse constraint f is π_f . Then the pricing problem of the column generation algorithm is as following:

$$\max_A R(A) - \sum_f \pi_f Q_f(A) - \pi. \quad (51)$$

The pricing problem is an assortment optimization with modified revenue $r'_j = r_j - \sum_f \pi_f B_{jf}$. We define a set of new variables x_j . $x_j = 1$ if product j is in the assortment and 0 otherwise. Problem (51) can be reformulated as following:

$$\max_x : \frac{\sum_{j \in \mathcal{J}} w_j r'_j x_j}{\sum_{j \in \mathcal{J}} w_j + v_0} \quad (52)$$

$$\text{s.t. } \sum_{j \in J_m} x_j = 1, \quad \forall m \quad (53)$$

$$x_j \in \{0, 1\}, \quad \forall j \quad (54)$$

The objective function (52) is the same as the revenue of an assortment under MNL choice model. The constraints (53) are totally unimodular. Thus, the algorithm in [14] can be used to solve the problem.

A more efficient way to solve the deterministic approximation problem is through the SBLP formulation. Let x_j denote the sales of product j . Let x_0 denote the number of customers that choose the no-purchase option. The SBLP formulation

with the 100% buy-down effects is as below.

$$\max_x : \sum_{j \in \mathcal{J}} r_j x_j \quad (55)$$

$$\text{s.t.} \sum_{j \in \mathcal{J}} x_j + x_0 = \lambda T \quad (56)$$

$$\sum_{j \in J} B_{jf} x_j \leq b_f, \quad \forall f \quad (57)$$

$$\sum_{j \in J_m} \frac{x_j}{w_j} \leq \frac{x_0}{v_0}, \quad \forall m \quad (58)$$

$$x_j \geq 0, \quad \forall j \quad (59)$$

Constraint (56) makes sure that each customer either buys some product or leave the system without buying anything. Constraint (58) guarantees that the no-purchase option is always available.

As a simplified version of Algorithm 1, Algorithm 2 below converts a feasible SBLP solution to a CDLP solution that leads to the same expected sales and revenue.

Algorithm 2 Algorithm of converting an SBLP solution to a CDLP solution

Input:

An SBLP solution \mathbf{x}
Set of products \mathcal{J} and its partition J_1, J_2, \dots, J_M
Attractiveness of product j : w_j , for all $j \in \mathcal{J}$
Attractiveness of the no-purchase alternative: v_0 .
Total arrival: $\lambda \cdot T$

Initialize:

Set $u(A) = 0$ for all $A \subset \mathcal{J}$ and $A \neq \emptyset$
Set $x = \mathbf{x}$
Set $i = 0$

- 1: **if** $x_j = 0$ for all $j \in \mathcal{J}$ **then**
- 2: Set $u(\emptyset) = 1 - \sum_{i'=1}^i u(A_{i'})$. Output u as the CDLP solution.
- 3: **for** $m = 1, \dots, M$ **do**
- 4: **if** $x_j = 0$ for all $j \in J_m$ **then**
- 5: Define $Y_m = -\infty$
- 6: **else**
- 7: Pick any $j(m) \in J_m$ such that $x_{j(m)} > 0$. Define

$$Y_m = \frac{x_{j(m)}}{w_{j(m)}}$$

- 8: Compute $I = \min \{Y_m : Y_m > 0\}$
- 9: Define

$$A_{i+1} = \bigcup_{m: Y_m > 0} \{j(m)\}$$

- 10: Define

$$u(A_{i+1}) = \frac{I (\sum_{m: Y_m > 0} w_{j(m)} + v_0)}{\lambda T}$$

- 11: **for** $m = 1, \dots, M$ **do**
- 12: **if** $Y_m > 0$ **then**
- 13: Update $x_{j(m)}$ as

$$x_{j(m)} = x_{j(m)} - I \cdot w_{j(m)}$$

- 14: Update $i = i + 1$. Go back to step 1
-

3.5 Assortment Optimization with Buy-down Effects and Random No-purchase

All choice models used in the revenue management literature have one special alternative called the no-purchase alternative. The no-purchase alternative aggregates all alternatives that are not from the company of interest. It includes the products

from competitors and the alternative of leaving the system without buying any thing. In many revenue management models with customer choice behavior, including the models we have discussed so far in this thesis, the no-purchase alternative is assumed to be known and constant. This is a strong assumption, especially considering that the no-purchase alternative includes all products from competitors. In practice, one has to admit that it is difficult to know what the competitors will offer in advance. The best a company can do is to have a probabilistic estimation of its competitors' offerings. In this case, the no-purchase alternative in the choice model will be random. Also, in practice, different customers may have different sets of alternatives as no-purchase. If the company cannot offer personalized assortments to customers, as is the case in practice, a better way to handle the inhomogeneity of customers' unknown no-purchase alternative is to view the no-purchase alternative as a random variable.

In this section, we study the assortment optimization problem with buy-down behavior and random no-purchase alternative. Determining the optimal assortment is an important problem in retailing or other industries where inventory is sufficient or managed through some other system. When the dynamic assortment optimization problem is approximated through a CDLP formulation, the subproblem of the column generation algorithm is actually an assortment optimization problem.

3.5.1 Notation and the assortment optimization problem

Let $\mathcal{J} = \{1, 2, \dots, J\}$ be the set of products the company has to offer. The revenue of product j is r_j . The collection of sets J_1, J_2, \dots, J_M form a partition of \mathcal{J} . Let $J(j)$ denote the subset that product j belongs to. Within each subset, if the assortment is not empty, the cheapest available product will receive the buy-down effects. When product j is the cheapest available among $J(j)$, its attractiveness is w_j , otherwise the attractiveness is v_j . We assume $w_j \geq v_j$. Let $I(j, A) = 1$ if product j is the

cheapest available among $J(j)$ when the assortment is A . Otherwise, $I(j, A) = 0$. The no-purchase alternative has L possible attractivenesses. The attractiveness is equal to v_{0l} with probability \mathbb{P}_l .

Given the notation, when the assortment is A , product j is chosen with probability equal to:

$$P_{j:A} = \sum_{l=1}^L \frac{I(j, A)w_j + (1 - I(j, A))v_j}{\sum_{j' \in A} [I(j', A)w_{j'} + (1 - I(j', A))v_{j'}] + v_{0l}} \mathbb{P}_l. \quad (60)$$

The corresponding assortment optimization is as following.

$$\max_{A \in 2^{\mathcal{J}}} : \sum_{j \in A} P_{j:A} r_j \quad (61)$$

In the remaining of the section, we first study the (static) assortment optimization problem when there is 100% buy-down effects. We develop a polynomial time algorithm to solve the assortment optimization problem. The component of the algorithm is to convert the original multi-dimensional optimization problem to a single-dimensional problem by heavily using the complementary slackness. The idea was first developed in [47], where it was used to solve a nonlinear nonseparable continuous knapsack problem. We show that the idea can also be adapted to solve the assortment optimization problem where the decision variables are actually discrete. We also show that the final single-dimensional problem can be solved within constant time. When there is general buy-down effects, we show that the optimal assortment is nested by revenue within each subset J_m . With the nesting property, we show that the assortment optimization problem with general buy-down effects and random no-purchase alternative can be reduced to an assortment optimization problem with 100% buy-down effects and random no-purchase alternative. Thus, the assortment optimization problem with general buy-down effects can also be solved within polynomial time.

3.5.2 Assortment optimization with 100% buy-down effects and random no-purchase alternative

We reformulate the assortment optimization problem with 100% buy-down effects and random no-purchase alternative. Let $x_j = 1$ if product j is included in the assortment and 0 otherwise. With such definition, given any binary vector x , we can quickly determine the corresponding assortment and denote it as $A(x)$. Since only the cheapest product within each J_m will be considered by the customers due to the 100% buy-down effects, we can impose the requirement that exactly one product from each J_m is set to be 1. To accommodate the case that there is no product from J_m is selected, we can introduce one dummy product to J_m . The dummy product has 0 attractiveness and 0 revenue. Thus, including the dummy product will not affect the actual choice probability and revenue. With such setup, the assortment optimization problem with 100% buy-down and random no-purchase alternative problem can be modeled as following:

$$\begin{aligned}
& \max_x : \sum_{l=1}^L \frac{\sum_{j \in \mathcal{J}} w_j r_j x_j}{\sum_{j \in \mathcal{J}} w_j x_j + v_{0l}} \mathbb{P}_l \\
& \text{such that : } \sum_{j \in J_m} x_j = 1, \quad \forall m \\
& x_j \in \{0, 1\}, \quad \forall j \in \mathcal{J}
\end{aligned} \tag{62}$$

Problem (62) is a nonlinear integer programming problem. If we remove the integer constraints, the optimization problem can be relaxed as (63) below. Note $x_j \leq 1$ is not required since the unit constraints will guarantee each x_j be bounded by 1.

$$\begin{aligned}
& \max_x : \sum_{l=1}^L \frac{\sum_{j \in \mathcal{J}} w_j r_j x_j}{\sum_{j \in \mathcal{J}} w_j x_j + v_{0l}} \mathbb{P}_l \\
& \text{such that : } \sum_{j \in J_m} x_j = 1, \quad \forall m \\
& x_j \geq 0, \quad \forall j \in \mathcal{J}
\end{aligned} \tag{63}$$

Theorem 9. *There always exists an optimal solution to problem (63) that is integral.*

Theorem 9 asserts that we can remove the integer constraints without introducing any optimality gap. Proof of Theorem 9 is postponed to the Appendix.

Since the logarithm function is a monotonically increasing function and the objective value of (63) is nonnegative for every feasible solution, the following problem, which is obtained by transforming the objective function through the logarithm function, has the same optimal solution as (63).

$$\begin{aligned}
Z^* := \max_x & : \log \left(\sum_{j \in \mathcal{J}} w_j r_j x_j \right) + \log \left(\sum_{l=1}^L \frac{\mathbb{P}_l}{\sum_{j \in \mathcal{J}} w_j x_j + v_{0l}} \right) \\
\text{such that : } & \sum_{j \in J_m} x_j = 1, \quad \forall m \\
& x_j \geq 0, \quad \forall j
\end{aligned} \tag{64}$$

We define a parametric function $f(\cdot) : \mathbb{R}_+ \rightarrow \mathbb{R}$ based on the sum $\sum_{j \in \mathcal{J}} w_j x_j$ as following:

$$\begin{aligned}
f(s) = \max_x & : \log \left(\sum_{j \in \mathcal{J}} w_j r_j x_j \right) + \log \left(\sum_{l=1}^L \frac{\mathbb{P}_l}{s + v_{0l}} \right) \\
\text{such that : } & \sum_{j \in J_m} x_j = 1, \quad \forall m \\
& \sum_{j \in \mathcal{J}} w_j x_j = s \\
& x_j \geq 0, \quad \forall j \in \mathcal{J}
\end{aligned} \tag{65}$$

The domain of s is $\left[\underline{s} := \sum_{m=1}^M \min_{j \in J_m} w_j, \bar{s} := \sum_{m=1}^M \max_{j \in J_m} w_j \right]$.

Theorem 10. $\max_{s \in [\underline{s}, \bar{s}]} f(s) = Z^*$.

Proof of Theorem 10. Suppose x^* is an optimal solution that leads to the optimal value Z^* in (64). Let $s^* = \sum_{j \in \mathcal{J}} w_j x_j^*$. Then $s^* \in [\underline{s}, \bar{s}]$ and x^* is a feasible solution to problem $f(s^*)$. Thus, we have

$$\max_{s \in [\underline{s}, \bar{s}]} f(s) \geq f(s^*) = z^* \tag{66}$$

On the other hand, suppose $f(s)$ is maximized at $s^* \in [\underline{s}, \bar{s}]$ and the corresponding solution to $f(s')$ is x' . Clearly, x' is also a feasible solution to problem (64) and the corresponding objective value is equal to $f(s')$. Thus:

$$Z^* \geq Z(x') = \max_{x \in [\underline{s}, \bar{s}]} f(s) \quad (67)$$

Combining (66) and (67), Theorem 10 is proved. \square

For a given s , the second term of the objective function of (65) is constant with respect to x . Problem (65) then has the same optimal solution as the following problem. This is also because the logarithm function is a monotonically increasing function.

$$\begin{aligned} \max_x : & \sum_{j \in \mathcal{J}} w_j r_j x_j \\ \text{such that : } & \sum_{j \in J_m} x_j = 1, \quad \forall m \\ & \sum_{j \in \mathcal{J}} w_j x_j = s \\ & x_j \geq 0, \quad \forall j \in \mathcal{J} \end{aligned} \quad (68)$$

The dual of problem (68) is as following:

$$\min_{\pi, \mu} : \sum_{m=1}^M \pi_m + s\mu \quad (69)$$

$$\text{such that : } \pi_m + w_j \mu \geq w_j r_j, \quad \forall m, \forall j \in J_m$$

where π_m is the dual variable corresponding to the m^{th} unit constraint and μ is the dual variable corresponding to the summation constraint.

Problem (68) has bounded feasible region, thus it has optimal solution. Also, since problem (68) has $m + 1$ equality constraints, there exists an optimal solution with $m + 1$ basic variables and only these basic variables can be positive. The m^{th} unit constraint implies that there is at least one positive x_j in each subset J_m . This gives us m basic variables, one from each J_m . Thus, the basic variable solution contains exactly two basic variables from the same subset.

Suppose for an optimal basic solution to (68), J_m contains two basic variables and $j_1, j_2 \in J_m$ are the two basic variables, specifically. According to the strict complementary slackness condition,

$$\pi_m + w_j \mu - w_j r_j = 0, \quad j = j_1, j_2. \quad (70)$$

From the system of equations (70), we can solve for μ and π_m . Once μ is determined, problem (69) can be solved and the optimal value of other π 's are $\pi_{m'} = \max_{j \in J_{m'}} w_j r_j - \mu w_j$, for all $m' \neq m$.

Given the optimal dual solution (μ, π) , according to the complementary slackness condition, in the optimal solution to problem (68), $x_j = 0$ if $\pi_{J(j)} + w_j \mu - w_j r_j > 0$. Let

$$\mathcal{S} = \{j : \pi_{J(j)} + w_j \mu - w_j r_j = 0\}. \quad (71)$$

Given the optimal dual solution (μ, π) , in the optimal solution to problem (68), only x_j with $j \in \mathcal{S}$ may be positive. All other x_j 's with $j \notin \mathcal{S}$ are equal to 0. Since The optimal solution to problem (68) is also optimal to problem (65), problem (65) can be expressed as

$$\begin{aligned} f_{\mu, \pi}(s) &= \log \left(\sum_{j \in \mathcal{S}} w_j r_j x_j \right) + \log \left(\sum_{l=1}^L \frac{\mathbb{P}_l}{s + v_{0l}} \right) \\ &= \log \left(\sum_{j \in \mathcal{S}} (\pi_{J(j)} + w_j \mu) x_j \right) + \log \left(\sum_{l=1}^L \frac{\mathbb{P}_l}{s + v_{0l}} \right) \\ &= \log \left(\sum_m \pi_m \sum_{j: j \in J_m \cap \mathcal{S}} x_j + \mu \sum_{j \in \mathcal{S}} w_j x_j \right) + \log \left(\sum_{l=1}^L \frac{\mathbb{P}_l}{s + v_{0l}} \right) \\ &= \log \left(\sum_m \pi_m + \mu s \right) + \log \left(\sum_{l=1}^L \frac{\mathbb{P}_l}{s + v_{0l}} \right), \end{aligned} \quad (72)$$

with $s = \sum_{j \in \mathcal{S}} w_j x_j$ and $x_j \geq 0$ for $j \in \mathcal{S}$.

The last equality holds since $\sum_{j \in J_m} x_j = 1$ and all x_j with $j \notin \mathcal{S}$ is equal to 0.

Note that once we determine the subset J_m as the only subset that contains two basic variables and determine which two products in J_m are the basic variables, the

corresponding dual optimal solution (μ, π) are determined. They are independent of the value of s . Thus, to optimize $f(s)$, we can optimize $f_{\mu, \pi}(s)$ over the domain of s for each (μ, π) pairs. We then compare the optimal value for all (μ, π) and pick the largest value. There are in total $\sum_m \frac{|J_m|(|J_m|-1)}{2}$ pairs of (μ, π) . Thus, if we can maximize $f_{\mu, \pi}(s)$ polynomially, the original assortment optimization problem can also be solved polynomially.

Theorem 11. *The function defined as (72) with feasible domain $[s_1, s_2]$ achieves its maximum either at s_1 or at s_2 .*

Theorem 11 says that for each given (μ, π) , to maximize $f_{\mu, \pi}(s)$, we only need to consider the smallest and largest possible value of s . Using this nice property, we develop Algorithm 3 to solve the assortment optimization problem with 100% buy-down effects and random no-purchase. The computational complexity of Algorithm 3 is $\mathcal{O}(MJ^2)$, where M is the number of subsets and J is the number of products. Proof of Theorem 11 is in Appendix.

Algorithm 3 Assortment Optimization Algorithm with 100% Buy-down and Random No-purchase

```

1: Set  $R = 0$ ,  $x = \mathbf{0}$ 
2: for  $m = 1, \dots, M$  do
3:   for  $j, j' \in J_m$  with  $j \neq j'$  do
4:     Solve for  $\mu$  and  $\pi_m$  from the two equations (70) for  $j$  and  $j'$ 
5:     procedure CHECK  $J_m$ 
6:       for  $k \in J_m$  and  $k \neq j, k \neq j'$  do
7:         if  $\pi_m + w_k \mu - w_k r_k < 0$  then
8:           continue to Step 3 with next pair of  $j$  and  $j'$ 
9:     for  $m' = 1, \dots, M$  do
10:       $\pi_{m'} = \min_{k \in J_{m'}} w_k r_k - w_k \mu$ 
11:       $\underline{w}_{m'} = \min\{w_k : w_k r_k - w_k \mu = \pi_{m'}, k \in J_{m'}\}$ 
12:       $\underline{x}_{m'} = \arg \min\{w_k : w_k r_k - w_k \mu = \pi_{m'}, k \in J_{m'}\}$ 
13:       $\bar{w}_{m'} = \max\{w_k : w_k r_k - w_k \mu = \pi_{m'}, k \in J_{m'}\}$ 
14:       $\bar{x}_{m'} = \arg \max\{w_k : w_k r_k - w_k \mu = \pi_{m'}, k \in J_{m'}\}$ 
15:       $\underline{w} = \sum_{m'} \underline{w}_{m'}$ ,  $\bar{w} = \sum_{m'} \bar{w}_{m'}$ 
16:      if  $f_{\mu, \pi}(\underline{w}) > f_{\mu, \pi}(\bar{w})$  then
17:         $R(m, j, j') = f_{\mu, \pi}(\underline{w})$ ;  $x' = \underline{x}$ 
18:      else
19:
20:         $R(m, j, j') = f_{\mu, \pi}(\bar{w})$ ;  $x' = \bar{x}$ 
21:      if  $R(m, j, j') > R$  then
22:        Set  $R = R(m, j, j')$ ,  $x = x'$ 
23: Output  $x$  as the optimal assortment.  $x_m$  is the index of the product selected from subset  $J_m$ .

```

3.5.3 Assortment optimization with general buy-down effects and random no-purchase

When the buy-down effects is 100%, the assortment optimization problem (61) essentially needs to determine which **one** of the products in J_m should be included in the assortment, for each subset $J_m \in \mathcal{J}$. This is so because even if we include two or more products in J_m , only the lowest available one will be considered by the customers. However, this will not be the case when the buy-down effects are not 100%. Multiple products from J_m can be included in the assortment since the more expensive ones will still be considered. In this sense, the assortment optimization problem with general buy-down effects is more difficult than the one with 100% buy-down effects. To

solve the problem, we will first show that the optimal solution to problem (61) with general buy-down effects has some nesting property.

Theorem 12. *There exists an optimal solution A^* to problem (61) that is nested by revenue within each subset. That is, if $j \in A^*$, $r_{j'} \geq r_j$, and $J(j') = J(j)$, then $j' \in A^*$.*

The proof of Theorem 12 is long and complicated. We postpone it to the Appendix for clarity. Given that the optimal assortment is nested in revenue within each subset J_m , we only need to determine the cheapest product to include in each J_m . Thus, we can reduce the assortment optimization problem with general buy-down effects to the one with 100% buy-down effects. The reduction algorithm can be stated as following.

Algorithm 4 Algorithm of reducing general buy-down effects to 100% buy-down effects

Input:

- Set of products \mathcal{J} and its partition J_1, J_2, \dots, J_M
 - Attractiveness of product j : v_j and w_j , for all $j \in \mathcal{J}$
 - Revenue of product j : r_j , for all $j \in \mathcal{J}$
 - 1: **for** $m = 1, \dots, M$ **do**
 - 2: **for** $j = 1, \dots, |J_m|$ **do**
 - 3: Let $A_{m,j}$ be the set of j most expensive products in J_m
 - 4: Let $w'_j = \sum_{j' \in A_{m,j}} [w_{j'} I(j', A_{m,j}) + v_{j'} (1 - I(j', A_{m,j}))]$
 - 5: Let $r'_j = \frac{\sum_{j' \in A_{m,j}} r_{j'} [w_{j'} I(j', A_{m,j}) + v_{j'} (1 - I(j', A_{m,j}))]}{w'_j}$
-

The idea of the reduction is to view all products in a nesting cut as an aggregate product. The fact that we can only choose one cut in each J_m is equivalent to saying that we can only choose one aggregate product from each J_m . r'_j and w'_j define a product in the 100% buy-down case. The partition of \mathcal{J} in the reduced problem is the same as before. We then can use Algorithm 3 to obtain the optimal assortment.

3.6 Conclusion

In this chapter, we study the revenue management problem with buy-down effects. We first give a very compact SBLP formulation for the dynamic assortment optimization problem under customer choice behavior with buy-down effects. In the SBLP formulation, both the number of constraints and the number of decision variables are at most quadratic in the problem input. We also give an efficient algorithm to convert an SBLP solution to a CDLP solution. Thus, the SBLP formulation is strictly dominating its CDLP counterpart. We then discuss the static assortment optimization problem under customer choice behavior with buy-down effects and random no-purchase alternative. The problem is of great practical importance since the alternatives from the competitors are part of the no-purchase alternative and they are rarely known in advance with certain. We develop efficient polynomial algorithm to solve the assortment optimization problem with 100% buy-down effects or general buy-down effects. The static assortment optimization problem can be used when the resource capacity is sufficient or in the subproblem of the column generation algorithm for the CDLP formulation. For industries where the buy-down effects are prominent, our treatment provides an economic formulation. Also, the solutions from our formulation and algorithm can be implemented easily.

APPENDIX A

PROOFS FOR CHAPTER 3

A.1 Proof of Theorem 9

The following lemma is useful when proving Theorem 9.

Lemma 11. *Consider a function $f : [-\delta, \delta] \rightarrow \mathbb{R}$ defined as following:*

$$f(x) = \sum_{l=1}^L \mathbb{P}_l \frac{cx + a}{bx + d_l},$$

where $c \geq 0, a \geq 0, \mathbb{P}_l \geq 0$ and $d_l > 0$ for every l , and $bx + d_l > 0$ for every $x \in [-\delta, \delta]$ and every l . Let $f'(x)$ denote the derivative of f at x . If $f'(0) \geq 0$, then $f(x) \geq 0$ for any $x \in [0, \delta]$.

Proof of Lemma 11. We have

$$f'(x) = \sum_{l=1}^L \mathbb{P}_l \frac{cd_l - ab}{(bx + d_l)^2}. \quad (73)$$

The denominator term in every summand of the derivative is always positive. When $b < 0$, since $\mathbb{P}_l > 0$ for every l , $c \geq 0$ and $a > 0$, the numerator term in every summand of the derivative is also nonnegative. Thus, when $b < 0$, $f'(x) \geq 0$ for every $x \in [-\delta, \delta]$. We now focus on the case when $b \geq 0$.

When $b \geq 0$, the numerator term $cd_l - ab \geq 0$ if $d_l \geq \frac{ab}{c}$ and $cd_l - ab < 0$ otherwise. Since the denominator of the summand is always positive, the sign of each summand is solely determined by the sign of its numerator. If $d_l \geq \frac{ab}{c}$ for every l , then every summand is nonnegative for all $x \in [-\delta, \delta]$. We now assume for some l 's, $d_l \geq \frac{ab}{c}$, and for others, $d_l < \frac{ab}{c}$. Without loss of generality, we assume

$$d_1 \leq d_2 \leq \dots \leq d_{k-1} < \frac{ab}{c} \leq d_k \leq \dots \leq d_L.$$

Given $f'(0) \geq 0$, we have

$$\sum_l \mathbb{P}_l \frac{cd_l - ab}{d_l^2} = \sum_{l=1}^{k-1} \mathbb{P}_l \frac{cd_l - ab}{d_l^2} + \sum_{l=k}^L \mathbb{P}_l \frac{cd_l - ab}{d_l^2} \geq 0 \quad (74)$$

$$\Rightarrow \sum_{l=k}^L \mathbb{P}_l \frac{cd_l - ab}{d_l^2} \geq \sum_{l=1}^{k-1} \mathbb{P}_l \frac{ab - cd_l}{d_l^2} \quad (75)$$

Note the summands at the left hand side and the right hand side are both nonnegative.

For $x \in [0, \delta]$

$$\begin{aligned} f'(x) &= \sum_{l=k}^L \mathbb{P}_l \frac{cd_l - ab}{(d_l + x)^2} - \sum_{l=1}^{k-1} \mathbb{P}_l \frac{ab - cd_l}{(d_l + x)^2} \\ &= \sum_{l=k}^L \mathbb{P}_l \frac{cd_l - ab}{d_l^2} \frac{d_l^2}{(d_l + x)^2} - \sum_{l=1}^{k-1} \mathbb{P}_l \frac{ab - cd_l}{d_l^2} \frac{d_l^2}{(d_l + x)^2} \\ &\geq \sum_{l=k}^L \mathbb{P}_l \frac{cd_l - ab}{d_l^2} \frac{d_k^2}{(d_k + x)^2} - \sum_{l=1}^{k-1} \mathbb{P}_l \frac{ab - cd_l}{d_l^2} \frac{d_k^2}{(d_k + x)^2} \\ &= f'(0) \frac{d_k^2}{(d_k + x)^2} \geq 0 \end{aligned} \quad (76)$$

Equality (76) because $\frac{d}{d+x} > \frac{d'}{d'+x} \geq 0$ as long as $d > d' \geq 0$ and $x > 0$.

Thus, in both cases, when $b < 0$ and when $b \geq 0$, we have $f'(x) \geq 0$ for $x \in [0, \delta]$ if $f'(0) \geq 0$. \square

The function in Lemma 11 has a nice property that it will be nondecreasing from any point onward as long as at that point the function is nondecreasing. We will show that the objective function of problem (63) is of this type.

Proof of Theorem 9. We prove the theorem by construction. Suppose an optimal solution x^* to problem (64) contains fractional values. Without loss of generality, we assume x_1^* and x_2^* with $1, 2 \in J_1$ are fractional. Also, we assume $w_1 r_1 \geq w_2 r_2$. We define function $f(z)$ as below:

$$f(z) = \sum_{l=1}^L \frac{\sum_{j=1}^J w_j r_j x_j^* + (w_1 r_1 - w_2 r_2)z}{\sum_{j=1}^J w_j x_j^* + v_{0l} + (w_1 - w_2)z} \mathbb{P}_l. \quad (77)$$

The domain of $f(z)$ is $[-x_1^*, x_2^*]$. $f(z)$ is the objective value if we shift z unit from x_2^* to x_1^* . By assumption, $c := w_1 r_1 - w_2 r_2 \geq 0$. We also have $d_l := \sum_{j \in J} w_j x_j^* + v_{0l} > 0$ for every l , and $\sum_{j \in J} w_j x_j^* + v_{0l} + (w_1 + w_2)z \geq v_{0l} > 0$ for every l and $z \in [-x_1^*, x_2^*]$. $\mathbb{P}_l \geq 0$ for every l and $a := \sum_{j=1}^J w_j r_j x_j^* \geq 0$ are obvious. Thus, function f satisfies all the requirements in Lemma 11. Since x^* (corresponding $z = 0$) is an optimal solution to problem (64) and 0 is an interior point in the domain of function f , we must have $f'(0) = 0$. Thus, according to Lemma 11, $f'(z) \geq 0$ for any $z \in [0, x_2^*]$. This means $f(x_2^*) \geq f(0)$. Thus shifting all of x_2^* to x_1^* leads to a solution that is also optimal. The new solution has one less fractional components. Repeating this process, we can remove all the fractional components and obtain an integral optimal solution. This proves the theorem. \square

A.2 Proof of Theorem 11

Lemma 12. Suppose $f_{\mu,\pi}(s) = \log(\pi + \mu s) + \log\left(\sum_{l=1}^L \frac{\mathbb{P}_l}{s + v_{0l}}\right)$ where v_{0l} and \mathbb{P}_l are positive for every l . The feasible region of f is $[\underline{s}, \bar{s}] \subset \mathbb{R}_{++}$. If $f'_{\mu,\pi}(s_0) = 0$ for some $s_0 \in (\underline{s}, \bar{s})$, then $f'_{\mu,\pi}(s) > 0$ for any $s \in (s_0, \bar{s}]$.

Proof of Lemma 12. The derivative of $f_{\mu,\pi}(s)$ is as following:

$$\begin{aligned} f'_{\mu,\pi}(s) &= \frac{\mu}{\pi + \mu s} - \frac{1}{\sum_l \frac{\mathbb{P}_l}{s + v_{0l}}} \sum_l \frac{\mathbb{P}_l}{(s + v_{0l})^2} \\ &= \frac{\sum_l \frac{\mathbb{P}_l}{(s + v_{0l})^2} \frac{v_{0l} - \frac{\pi}{\mu}}{s + \frac{\pi}{\mu}}}{\sum_l \frac{\mathbb{P}_l}{s + v_{0l}}} \end{aligned}$$

If $f_{\mu,\pi}(s_0) = 0$, we have:

$$\frac{\sum_l \frac{\mathbb{P}_l}{(s_0 + v_{0l})^2} \frac{v_{0l} - \frac{\pi}{\mu}}{s_0 + \frac{\pi}{\mu}}}{\sum_l \frac{\mathbb{P}_l}{s_0 + v_{0l}}} = 0 \Rightarrow \sum_l \frac{\mathbb{P}_l}{(s_0 + v_{0l})^2} \frac{v_{0l} - \frac{\pi}{\mu}}{s_0 + \frac{\pi}{\mu}} = 0. \quad (78)$$

Since $\sum_l \frac{\mathbb{P}_l}{s + v_{0l}} > 0$ for any $s \in [\underline{s}, \bar{s}]$, to prove $f'_{\mu,\pi}(s) > 0$ for $s \in (s_0, \bar{s}]$, we only need to show

$$\sum_l \frac{\mathbb{P}_l}{(s + v_{0l})^2} \frac{v_{0l} - \frac{\pi}{\mu}}{s + \frac{\pi}{\mu}} > 0. \quad (79)$$

Note that the l^{th} summand in (78) and (79) have the same sign. Each of them will be positive if and only if $v_{0l} > \frac{\pi}{\mu}$. Without loss of generality, suppose:

$$v_{01} \leq \cdots \leq v_{0l^*} < \frac{\pi}{\mu} \leq v_{0,l^*+1} \leq \cdots \leq v_{0L}.$$

We can relax the sum of the positive terms on the left hand side of (79) as following:

$$\begin{aligned} & \sum_{l=l^*+1}^L \frac{\mathbb{P}_l}{(s+v_{0l})^2} \frac{v_{0l} - \frac{\pi}{\mu}}{s + \frac{\pi}{\mu}} \\ &= \sum_{l=l^*+1}^L \frac{\mathbb{P}_l \left(v_{0l} - \frac{\pi}{\mu} \right)}{\left(s_0 + \frac{\pi}{\mu} \right) (s_0 + v_{0l})^2} \frac{\left(s_0 + \frac{\pi}{\mu} \right) (s_0 + v_{0l})^2}{\left(s + \frac{\pi}{\mu} \right) (s + v_{0l})^2} \\ &> \sum_{l=l^*+1}^L \frac{\mathbb{P}_l \left(v_{0l} - \frac{\pi}{\mu} \right)}{\left(s_0 + \frac{\pi}{\mu} \right) (s_0 + v_{0l})^2} \frac{\left(s_0 + \frac{\pi}{\mu} \right) (s_0 + v_{0l^*})^2}{\left(s + \frac{\pi}{\mu} \right) (s + v_{0l^*})^2} \end{aligned} \quad (80)$$

The last inequality is because $\frac{a+x}{b+x}$ is strictly increasing with x if $b > a > 0$ and $x > 0$.

Similarly, we relax the sum of the negative terms on the left hand side of (79) as following:

$$\begin{aligned} & \sum_{l=1}^{l^*} \frac{\mathbb{P}_l}{(s+v_{0l})^2} \frac{v_{0l} - \frac{\pi}{\mu}}{s + \frac{\pi}{\mu}} \\ &= \sum_{l=1}^{l^*} \frac{\mathbb{P}_l \left(v_{0l} - \frac{\pi}{\mu} \right)}{\left(s_0 + \frac{\pi}{\mu} \right) (s_0 + v_{0l})^2} \frac{\left(s_0 + \frac{\pi}{\mu} \right) (s_0 + v_{0l})^2}{\left(s + \frac{\pi}{\mu} \right) (s + v_{0l})^2} \\ &> \sum_{l=1}^{l^*} \frac{\mathbb{P}_l \left(v_{0l} - \frac{\pi}{\mu} \right)}{\left(s_0 + \frac{\pi}{\mu} \right) (s_0 + v_{0l})^2} \frac{\left(s_0 + \frac{\pi}{\mu} \right) (s_0 + v_{0l^*})^2}{\left(s + \frac{\pi}{\mu} \right) (s + v_{0l^*})^2} \end{aligned} \quad (81)$$

Combining (80) and (81), we have

$$\sum_l \frac{\mathbb{P}_l}{(s+v_{0l})^2} \frac{v_{0l} - \frac{\pi}{\mu}}{s + \frac{\pi}{\mu}} > \frac{\left(s_0 + \frac{\pi}{\mu} \right) (s_0 + v_{0l^*})^2}{\left(s + \frac{\pi}{\mu} \right) (s + v_{0l^*})^2} \left(\sum_l \frac{\mathbb{P}_l}{(s_0 + v_{0l})^2} \frac{v_{0l} - \frac{\pi}{\mu}}{s_0 + \frac{\pi}{\mu}} \right) = 0.$$

Thus, inequality (79) is proved and we have $f'_{\mu,\pi}(s) \geq 0$ for $s \in [s_0, \bar{s}]$. \square

Proof of Theorem 11. We prove the theorem by contradiction. Suppose $f_{\mu,\pi}$ achieves its maximum at $s_0 \in (\underline{s}, \bar{s})$, then $f_{\mu,\pi}(s_0) = 0$. According to Lemma 12, $f'_{\mu,\pi}(s) > 0$ for $s \in (s_0, \bar{s}]$. Thus, $f_{\mu,\pi}(\bar{s}) > f_{\mu,\pi}(s_0)$. This contradicts the assumption that s_0 is a maximum solution. This proves the result. \square

A.3 Proof of Theorem 12

To prove Theorem 12, we consider another optimization problem which focuses on the assortment optimization within subset J_1 . For any $A \subset J_1$, let:

$$R(A) = \sum_{l=1}^L \frac{\sum_{j \in A} (I(j, A)w_j + (1 - I(j, A))v_j) r_j + v_f r_f}{\sum_{j \in A} (I(j, A)w_j + (1 - I(j, A))v_j) + v_f + v_{0l}} \mathbb{P}_l, \quad (82)$$

where f is some given assortment from subsets J_2, \dots, J_M . Its total attractiveness and weighted revenue are denoted as v_f and r_f , respectively. The optimization problem of focus is:

$$R^* = \max_{A \subset J_1} R(A). \quad (83)$$

For notational brevity, we also define the following:

$$R'(A) = \sum_{l=1}^L \frac{\sum_{j \in A} v_j r_j + v_f r_f}{\sum_{j \in A} v_j + v_f + v_{0l}} \mathbb{P}_l. \quad (84)$$

$$A_j = \{1, 2, \dots, j\} \quad (85)$$

$$r(A) = \frac{\sum_{j \in A} (I(j, A)w_j + (1 - I(j, A))v_j) r_j}{\sum_{j \in A} (I(j, A)w_j + (1 - I(j, A))v_j)} \quad (86)$$

$$r'(A) = \frac{\sum_{j \in A} v_j r_j}{\sum_{j \in A} v_j} \quad (87)$$

$$w(A) = \sum_{j \in A} (I(j, A)w_j + (1 - I(j, A))v_j) \quad (88)$$

$$v(A) = \sum_{j \in A} v_j \quad (89)$$

The following two mathematical lemmas are used when we prove properties of the optimal solution to Problem (83).

Lemma 13. *Given any $b, r, c, x, a_1, a_2, y_1$, and y_2 , all being positive numbers, if the following condition is satisfied*

$$y_1 \frac{b - ca_1}{(a_1 + r)(a_1 + x)} + y_2 \frac{b - ca_2}{(a_2 + r)(a_2 + x)} > 0, \quad (90)$$

then there exists $a' \in (0, b/c)$ and $y' > 0$ such that the following holds:

$$y_1 \frac{b - ca_1}{(a_1 + r)(a_1 + x)} + y_2 \frac{b - ca_2}{(a_2 + r)(a_2 + x)} \leq y' \frac{b - ca'}{(a' + r)(a' + x)}, \quad (91)$$

$$y_1 \frac{b - ca_1}{a_1(a_1 + r)(a_1 + x)} + y_2 \frac{b - ca_2}{a_2(a_2 + r)(a_2 + x)} \geq y' \frac{b - ca'}{a'(a' + r)(a' + x)}. \quad (92)$$

Proof. Without loss of generality, we assume $a_1 \leq a_2$. Because of inequality (90), we have $y_1 \frac{b - ca_1}{(a_1 + r)(a_1 + x)} > 0$. Thus, $a_1 < b/c$. Next, we discuss the case for $a_2 > b/c$ and the case $a_2 \leq b/c$ separately.

- Case 1: $a_2 \geq b/c$. We first observe that

$$\begin{aligned} & y_1 \frac{b - ca_1}{a_1(a_1 + r)(a_1 + x)} + y_2 \frac{b - ca_2}{a_2(a_2 + r)(a_2 + x)} \\ & \geq y_1 \frac{b - ca_1}{a_2(a_1 + r)(a_1 + x)} + y_2 \frac{b - ca_2}{a_2(a_2 + r)(a_2 + x)} > 0 \end{aligned}$$

Let $y' = y_1 + y_2$. Since $g(a) := \frac{b - ca}{a(a + r)(a + x)}$ is continuous and decreasing on interval $(0, b/c)$, there exists a unique $a' \in (a_1, b/c)$ such that

$$y' \frac{b - ca'}{a'(a' + r)(a' + x)} = y_1 \frac{b - ca_1}{a_1(a_1 + r)(a_1 + x)} + y_2 \frac{b - ca_2}{a_2(a_2 + r)(a_2 + x)}.$$

Thus, y' and a' selected as above satisfy Inequality (92). We now show that Inequality (91) is satisfied by y' and a' as well.

$$\begin{aligned} y' \frac{b - ca'}{(a' + r)(a' + x)} &= a' y' \frac{b - ca'}{a'(a' + r)(a' + x)} \\ &= a' \left(y_1 \frac{b - ca_1}{a_1(a_1 + r)(a_1 + x)} + y_2 \frac{b - ca_2}{a_2(a_2 + r)(a_2 + x)} \right) \\ &= \frac{a'}{a_1} y_1 \frac{b - ca_1}{(a_1 + r)(a_1 + x)} + \frac{a'}{a_2} y_2 \frac{b - ca_2}{a_2(a_2 + r)(a_2 + x)} \\ &\geq y_1 \frac{b - ca_1}{(a_1 + r)(a_1 + x)} + y_2 \frac{b - ca_2}{(a_2 + r)(a_2 + x)} \end{aligned}$$

The last inequality follows because $a' \in (a_1, a_2)$, $y_1 \frac{b - ca_1}{(a_1 + r)(a_1 + x)} > 0$, and

$$y_2 \frac{b - ca_2}{(a_2 + r)(a_2 + x)} \leq 0.$$

- Case 2: $a_2 < b/c$. Rearrange (91) and (92) we have the following:

$$y' \geq \frac{y_1 \frac{b-ca_1}{(a_1+r)(a_1+x)} + y_2 \frac{b-ca_2}{(a_2+r)(a_2+x)}}{\frac{b-ca'}{(a'+r)(a'+x)}}$$

$$y' \leq \frac{y_1 \frac{b-ca_1}{a_1(a_1+r)(a_1+x)} + y_2 \frac{b-ca_2}{a_2(a_2+r)(a_2+x)}}{\frac{b-ca'}{a'(a'+r)(a'+x)}}$$

One pair of feasible (a', y') can be found if we find some $a' \in (0, b/c)$ such that

$$\frac{y_1 \frac{b-ca_1}{(a_1+r)(a_1+x)} + y_2 \frac{b-ca_2}{(a_2+r)(a_2+x)}}{\frac{b-ca'}{(a'+r)(a'+x)}} \leq \frac{y_1 \frac{b-ca_1}{a_1(a_1+r)(a_1+x)} + y_2 \frac{b-ca_2}{a_2(a_2+r)(a_2+x)}}{\frac{b-ca'}{a'(a'+r)(a'+x)}}, \quad (93)$$

which is equivalent to

$$\frac{y_1 \frac{b-ca_1}{(a_1+r)(a_1+x)} + y_2 \frac{b-ca_2}{(a_2+r)(a_2+x)}}{y_1 \frac{b-ca_1}{a_1(a_1+r)(a_1+x)} + y_2 \frac{b-ca_2}{a_2(a_2+r)(a_2+x)}} \leq a'. \quad (94)$$

Once such an a' is found, then y' can be chosen as any number between the LHS and RHS of Inequality (93). With both the numerator and the denominator multiplied with $a_1 a_2$, the LHS of Inequality (94) can be rewritten as

$$\frac{\left(y_1 \frac{b-ca_1}{(a_1+r)(a_1+x)} a_2\right) a_1 + \left(y_2 \frac{b-ca_2}{(a_2+r)(a_2+x)} a_1\right) a_2}{y_1 \frac{b-ca_1}{(a_1+r)(a_1+x)} a_2 + y_2 \frac{b-ca_2}{(a_2+r)(a_2+x)} a_1}$$

which is a convex combination of a_1 and a_2 . Thus the LHS of Inequality (94) is between a_1 and a_2 . Thus, if we pick a' equal to the LHS of Inequality (94) and pick y' any number between LHS and RHS of Inequality (93), both (91) and (92) will be satisfied.

To conclude, we establish the results. □

Lemma 14. *Given b, r, c and x , all being positive, let a_1, a_2, \dots, a_L , and y_1, y_2, \dots, y_L be any positive real numbers. If $A := \sum_{l=1}^L y_l \frac{b-ca_l}{(a_l+r)(a_l+x)} > 0$, then*

$$B := \sum_{l=1}^L y_l \frac{b-ca_l}{a_l(a_l+r)(a_l+x)} > 0.$$

Proof. We prove the result by induction on L . For $L = 1$, the result trivially holds.

We assume the result holds for $L = k$ with $k \geq 1$. We next show that the result also

holds for $L = k + 1$. Without loss of generality, we assume $a_1 \leq a_2 \leq \cdots \leq a_{k+1}$.

Since $A > 0$, we have $y_1 \frac{b-ca_1}{(a_1+r)(a_1+x)} + y_2 \frac{b-ca_2}{(a_2+r)(a_2+x)} > 0$. According to Lemma 13, there exists $y' > 0$ and $a' > 0$ such that Inequality (91) and Inequality (92) hold.

Now a', a_3, \dots, a_{k+1} , and y', y_3, \dots, y_{k+1} can be viewed as a case when $L = k$. Since

$y' \frac{b-ca'}{(a'+r)(a'+x)} + \sum_{l=3}^L y_l \frac{b-ca_l}{(a_l+r)(a_l+x)} > 0$, according to the inductive assumption, we have

$$y' \frac{b-ca'}{(a'+r)(a'+x)} + \sum_{l=3}^L y_l \frac{b-ca_l}{a_l(a_l+r)(a_l+x)} > 0.$$

Since $B \geq y' \frac{b-ca'}{(a'+r)(a'+x)} + \sum_{l=3}^L y_l \frac{b-ca_l}{a_l(a_l+r)(a_l+x)}$, we establish the results for $L = k+1$. \square

Lemma 15. *Given b, r, c and x , all being positive, let a_1, a_2, \dots, a_L , and y_1, y_2, \dots, y_L be any positive real numbers. If $A := \sum_{l=1}^L y_l \frac{b-ca_l}{(a_l+r)(a_l+x)} > 0$, then*

$$C := \sum_{l=1}^L y_l \frac{b-ca_l}{(a_l)(a_l+x)} > 0.$$

Proof. $C - A = r \sum_{l=1}^L y_l \frac{b-ca_l}{a_l(a_l+r)(a_l+x)}$. According to Lemma 14, Given $A > 0$, $C - A > 0$. Thus, $C > 0$. \square

Lemma 16. *If $j_0 \geq 1$ satisfies $R(A_{j_0+1}) < R'(A_{j_0+1})$, then $R'(A_{j_0+1}) < R'(A_{j_0})$.*

Proof. According to assumption, $v_{j_0+1} \leq w_{j_0+1}$. Since $R(A_{j_0+1}) < R'(A_{j_0+1})$, we have

$$v_{j_0+1} < w_{j_0+1}.$$

$$\begin{aligned} & R(A_{j_0+1}) < R'(A_{j_0+1}) \\ \Leftrightarrow & \sum_{l=1}^L \frac{v(A_{j_0})r'(A_{j_0}) + w_{j_0+1}r_{j_0+1} + v_f r_f}{v(A_{j_0}) + w_{j_0+1} + v_f + v_{0l}} \mathbb{P}_l < \sum_{l=1}^L \frac{v(A_{j_0})r'(A_{j_0}) + v_{j_0+1}r_{j_0+1} + v_f r_f}{v(A_{j_0}) + v_{j_0+1} + v_f + v_{0l}} \mathbb{P}_l \\ \Rightarrow & \sum_{l=1}^L \mathbb{P}_l \frac{(w_{j_0+1} - v_{j_0+1})(v(A_{j_0})r'(A_{j_0}) + v_f r_f - r_{j_0+1}(v(A_{j_0}) + v_f + v_{0l}))}{(v(A_{j_0}) + v_{j_0+1} + v_f + v_{0l})(v(A_{j_0}) + w_{j_0+1} + v_f + v_{0l})} > 0 \\ \Rightarrow & \sum_{l=1}^L \mathbb{P}_l \frac{v(A_{j_0})r'(A_{j_0}) + v_f r_f - r_{j_0+1}(v(A_{j_0}) + v_f + v_{0l})}{(v(A_{j_0}) + v_{j_0+1} + v_f + v_{0l})(v(A_{j_0}) + w_{j_0+1} + v_f + v_{0l})} > 0 \end{aligned} \quad (95)$$

The difference between $R'(A_{j_0+1})$ and $R'(A_{j_0})$ can be expressed as following:

$$\begin{aligned}
& R'(A_{j_0}) - R'(A_{j_0+1}) \\
&= \sum_{l=1}^L \mathbb{P}_l \left(\frac{v(A_{j_0})r'(A_{j_0}) + v_f r_f}{v(A_{j_0}) + v_f + v_{0l}} - \frac{v(A_{j_0})r'(A_{j_0}) + v_{j_0+1}r_{j_0+1} + v_f r_f}{v(A_{j_0}) + v_{j_0+1} + v_f + v_{0l}} \right) \\
&= v_{j_0+1} \sum_{l=1}^L \mathbb{P}_l \frac{v(A_{j_0})r'(A_{j_0}) + v_f r_f - r_{j_0+1}(v(A_{j_0}) + v_f + v_{0l})}{(v(A_{j_0}) + v_f + v_{0l})(v(A_{j_0}) + v_{j_0+1} + v_f + v_{0l})}.
\end{aligned}$$

Viewing $v(A_{j_0}) + v_f + v_{0l}$ as a_l , v_{j_0+1} as x , w_{j_0+1} as r , r_{j_0+1} as c , $v(A_{j_0})r'(A_{j_0}) + v_f r_f$ as b , \mathbb{P}_l as y_l , and with Inequality (95), we can apply Lemma 15 and have $R'(A_{j_0}) - R'(A_{j_0+1}) > 0$. \square

Lemma 17. *If $j_0 \geq 1$ satisfies $R(A_{j_0+1}) < R'(A_{j_0+1})$, then there exists $j^* < j_0$ such that $R'(A_{j_0+1}) < R(A_{j^*})$.*

Proof. Let $J^* = \{j : j < j_0, R'(A_j) \leq R(A_j)\}$. J^* is not empty since $0 \in J^*$ (A_0 denotes the empty assortment). Let $j^* = \max J^*$. By definition of J^* and j^* , for any $j \in (j^*, j_0]$, $R(A_j) < R'(A_j)$. Due to Lemma 16, $R'(A_{j^*}) > R'(A_{j^*+1}) > \dots > R'(A_{j_0+1})$. By definition, we have $R(A_{j^*}) \geq R'(A_{j^*})$. Thus, we have $R'(A_{j_0+1}) < R(A_{j^*})$ and the result is proved. \square

Corollary 2. *If $j_0 \geq 1$ satisfies $R(A_{j_0+1}) < R'(A_{j_0+1})$, then $R'(A_{j_0+1}) < R^*$.*

Theorem 13. *For any $j \in \mathcal{J}$, $R'(A_j) \leq R^*$.*

Proof. If $R(A_j) < R'(A_j)$, according to Corollary 2, $R'(A_j) < R^*$. If $R(A_j) \geq R'(A_j)$, since $R^* \geq R(A_j)$, we have $R'(A_j) \leq R^*$. To conclude, we have $R'(A_j) \leq R^*$ for every j . \square

Lemma 18. *Suppose A is an optimal solution to Problem (82). If there exists $j \notin A$ such that $\underline{A} := \{j' \in A : r_{j'} < r_j\} \neq \emptyset$, then $A \cup \{j\}$ is also an optimal solution to Problem (82).*

Proof. We define $\bar{A} := \{j' \in A : r_{j'} \geq r_j\}$. To show $A \cup \{j\}$ is also an optimal solution, we only need to show $R(A \cup \{j\}) \geq R(A)$, which is equivalent to show:

$$\sum_{l=1}^L \mathbb{P}_l \frac{v(\bar{A})r'(\bar{A}) + v_j r_j + w(\underline{A})r(\underline{A}) + v_f r_f}{v(\bar{A}) + v_j + w(\underline{A}) + v_f + v_{0l}} \geq \sum_{l=1}^L \mathbb{P}_l \frac{v(\bar{A})r'(\bar{A}) + w(\underline{A})r(\underline{A}) + v_f r_f}{v(\bar{A}) + w(\underline{A}) + v_f + v_{0l}} \quad (96)$$

$$\begin{aligned} & \Leftrightarrow r_j \sum_{l=1}^L \mathbb{P}_l \frac{1}{v(\bar{A}) + v_j + w(\underline{A}) + v_f + v_{0l}} \\ & \geq \sum_{l=1}^L \mathbb{P}_l \frac{v(\bar{A})r'(\bar{A}) + w(\underline{A})r(\underline{A}) + v_f r_f}{(v(\bar{A}) + v_j + w(\underline{A}) + v_f + v_{0l})(v(\bar{A}) + w(\underline{A}) + v_f + v_{0l})} \end{aligned} \quad (97)$$

Since $r_j \geq r(\underline{A})$, Inequality (97) can be established if the following inequality holds:

$$\begin{aligned} & r(\underline{A}) \sum_{l=1}^L \mathbb{P}_l \frac{1}{v(\bar{A}) + v_j + w(\underline{A}) + v_f + v_{0l}} \\ & \geq \sum_{l=1}^L \mathbb{P}_l \frac{v(\bar{A})r'(\bar{A}) + w(\underline{A})r(\underline{A}) + v_f r_f}{(v(\bar{A}) + v_j + w(\underline{A}) + v_f + v_{0l})(v(\bar{A}) + w(\underline{A}) + v_f + v_{0l})}, \end{aligned}$$

which is equivalent to:

$$\begin{aligned} & r(\underline{A}) \sum_{l=1}^L \mathbb{P}_l \frac{v(\bar{A}) + v_f + v_{0l}}{(v(\bar{A}) + v_j + w(\underline{A}) + v_f + v_{0l})(v(\bar{A}) + w(\underline{A}) + v_f + v_{0l})} \\ & \geq \sum_{l=1}^L \mathbb{P}_l \frac{v(\bar{A})r'(\bar{A}) + v_f r_f}{(v(\bar{A}) + v_j + w(\underline{A}) + v_f + v_{0l})(v(\bar{A}) + w(\underline{A}) + v_f + v_{0l})}. \end{aligned} \quad (98)$$

On the other hand, since A is optimal, according to Theorem 13, $R(A) \geq R'(\bar{A})$.

Thus,

$$\begin{aligned} & \sum_{l=1}^L \mathbb{P}_l \frac{v(\bar{A})r'(\bar{A}) + w(\underline{A})r(\underline{A}) + v_f r_f}{v(\bar{A}) + w(\underline{A}) + v_f + v_{0l}} \geq \sum_{l=1}^L \mathbb{P}_l \frac{v(\bar{A})r'(\bar{A}) + v_f r_f}{v(\bar{A}) + v_f + v_{0l}} \\ & \Rightarrow \sum_{l=1}^L \mathbb{P}_l \frac{r(\underline{A})}{v(\bar{A}) + w(\underline{A}) + v_f + v_{0l}} \geq \sum_{l=1}^L \mathbb{P}_l \frac{v(\bar{A})r'(\bar{A}) + v_f r_f}{(v(\bar{A}) + w(\underline{A}) + v_f + v_{0l})(v(\bar{A}) + v_f + v_{0l})}. \end{aligned} \quad (99)$$

Inequality (98) can be established if the following holds:

$$\frac{\sum_{l=1}^L \mathbb{P}_l \frac{v(\bar{A})r'(\bar{A}) + v_f r_f}{(v(\bar{A}) + w(\underline{A}) + v_f + v_{0l})(v(\bar{A}) + v_f + v_{0l})}}{\sum_{l=1}^L \mathbb{P}_l \frac{1}{v(\bar{A}) + w(\underline{A}) + v_f + v_{0l}}} \geq \frac{\sum_{l=1}^L \mathbb{P}_l \frac{v(\bar{A})r'(\bar{A}) + v_f r_f}{(v(\bar{A}) + v_j + w(\underline{A}) + v_f + v_{0l})(v(\bar{A}) + w(\underline{A}) + v_f + v_{0l})}}{\sum_{l=1}^L \mathbb{P}_l \frac{v(\bar{A}) + v_f + v_{0l}}{(v(\bar{A}) + v_j + w(\underline{A}) + v_f + v_{0l})(v(\bar{A}) + w(\underline{A}) + v_f + v_{0l})}}, \quad (100)$$

which is equivalent to show:

$$\begin{aligned}
& \sum_{l=1}^L \sum_{l'=1}^L \frac{\mathbb{P}_l \mathbb{P}'_{l'} (v(\bar{A}) + v_f + v_{0l'}) / (v(\bar{A}) + v_f + v_{0l})}{(v(\bar{A}) + w(\underline{A}) + v_f + v_{0l})(v(\bar{A}) + v_j + w(\underline{A}) + v_f + v_{0l'})(v(\bar{A}) + w(\underline{A}) + v_f + v_{0l'})} \\
& \geq \sum_{l=1}^L \sum_{l'=1}^L \frac{\mathbb{P}_l \mathbb{P}'_{l'}}{(v(\bar{A}) + v_j + w(\underline{A}) + v_f + v_{0l})(v(\bar{A}) + w(\underline{A}) + v_f + v_{0l'})(v(\bar{A}) + w(\underline{A}) + v_f + v_{0l'})}.
\end{aligned} \tag{101}$$

For $l = l'$, the cross terms are canceled out on both sides. For any $l \neq l'$, we can simplify the cross terms by canceling the common term $(v(\bar{A}) + w(\underline{A}) + v_f + v_{0l})(v(\bar{A}) + w(\underline{A}) + v_f + v_{0l'})$ in the denominator and we equivalently show the following:

$$\begin{aligned}
& \frac{v(\bar{A}) + v_f + v_{0l'}}{v(\bar{A}) + v_f + v_{0l}} \frac{1}{(v(\bar{A}) + v_j + w(\underline{A}) + v_f + v_{0l'})} \\
& + \frac{v(\bar{A}) + v_f + v_{0l}}{v(\bar{A}) + v_f + v_{0l'}} \frac{1}{(v(\bar{A}) + v_j + w(\underline{A}) + v_f + v_{0l})} \\
& \geq \frac{1}{(v(\bar{A}) + v_j + w(\underline{A}) + v_f + v_{0l})} + \frac{1}{(v(\bar{A}) + v_j + w(\underline{A}) + v_f + v_{0l'})}. \\
& \Leftrightarrow \frac{v_{0l'} - v_{0l}}{(v(\bar{A}) + v_f + v_{0l})(v(\bar{A}) + v_j + w(\underline{A}) + v_f + v_{0l'})} \\
& + \frac{v_{0l} - v_{0l'}}{(v(\bar{A}) + v_f + v_{0l'})(v(\bar{A}) + v_j + w(\underline{A}) + v_f + v_{0l})} \geq 0 \\
& \Leftrightarrow (v_{0l} - v_{0l'})^2 (v_j + w(\underline{A})) \geq 0
\end{aligned}$$

The last inequality is straightforward. Thus, we have $R(A \cup \{j\}) \geq R(A)$ and $A \cup \{j\}$ is also an optimal solution to Problem (82). \square

Due to Lemma 18, Theorem 12 can be proved in a straightforward way.

REFERENCES

- [1] ALGERS, S. and BESER, M., “Modelling choice of flight and booking class—a study using stated preference and revealed preference data,” *International Journal of Services Technology and Management*, vol. 2, no. 1-2, pp. 28–45, 2001.
- [2] ANDERSSON, S.-E., “Passenger choice analysis for seat capacity control: A pilot project in scandinavian airlines,” *International Transactions in Operational Research*, vol. 5, no. 6, pp. 471–486, 1998.
- [3] BELOBABA, P., *Air travel demand and airline seat inventory management*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 1987.
- [4] BELOBABA, P. P., “Survey paper-airline yield management an overview of seat inventory control,” *Transportation Science*, vol. 21, no. 2, pp. 63–73, 1987.
- [5] BELOBABA, P. P., “Application of a probabilistic decision model to airline seat inventory control,” *Operations Research*, vol. 37, no. 2, pp. 183–197, 1989.
- [6] BEN-AKIVA, M. E. and LERMAN, S. R., *Discrete Choice Analysis: Theory and Application to Travel Demand*, vol. 9. Cambridge, MA: MIT press.
- [7] BERTSEKAS, D. and TSITSIKLIS, J., *Neuro-Dynamic Programming*. Belmont, MA: Athena Scientific, 1996.
- [8] BLANCHET, J., GALLEGO, G., and GOYAL, V., “A markov chain approximation to choice modeling.” Working paper, School of IEOR, Columbia University, 2013.
- [9] BOYD, E. A. and KALLESEN, R., “The science of revenue management when passengers purchase the lowest available fare,” *Journal of Revenue and Pricing Management*, vol. 3, no. 2, pp. 171–177, 2004.
- [10] BRONT, J. J. M., MÉNDEZ-DÍAZ, I., and VULCANO, G., “A column generation algorithm for choice-based network revenue management,” *Operations Research*, vol. 57, no. 3, pp. 769–784, 2009.
- [11] COOPER, W. L., HOMEM-DE MELLO, T., and KLEYWEGT, A. J., “Models of the spiral-down effect in revenue management,” *Operations Research*, vol. 54, no. 5, pp. 968–987, 2006.
- [12] CROSS, R. G., *Revenue Management: Hard-core Tactics for Market Domination*. New York, NY: Broadway Books, 1997.

- [13] DAI, J., DING, W., KLEYWEGT, A., WANG, X., and ZHANG, Y., “Choice based revenue management for parallel flights.” Working paper, School of Industrial and Systems Engineering, Georgia Tech, 2014.
- [14] DAVIS, J., GALLEG0, G., and TOPAL0GLU, H., “Assortment planning under the multinomial logit model with totally unimodular constraint structures.” Working paper, School of ORIE, Cornell University, 2013.
- [15] DAVIS, J. M., GALLEG0, G., and TOPAL0GLU, H., “Assortment optimization under variants of the nested logit model,” *Operations Research*, vol. 62, no. 2, pp. 250–273, 2014.
- [16] DAVIS, J. M., TOPAL0GLU, H., and WILLIAMSON, D. P., “Pricing problems under the nested logit model with a quality consistency constraint.” Working paper, School of ORIE, Cornell University, 2015.
- [17] DÉSIR, A. and GOYAL, V., “Near-optimal algorithms for capacity constrained assortment optimization.” Available at SSRN 2543309, 2014.
- [18] DÉSIR, A., GOYAL, V., SEGEV, D., and YE, C., “Capacity constrained assortment optimization under the markov chain based choice model,” *Operations Research*, *Forthcoming*, 2015.
- [19] FARIAS, V. F., JAGABATHULA, S., and SHAH, D., “A nonparametric approach to modeling choice with limited data,” *Management Science*, vol. 59, no. 2, pp. 305–322, 2013.
- [20] FELDMAN, J. and TOPAL0GLU, H., “Assortment optimization under the multinomial logit model with nested consideration sets.” Working paper, School of ORIE, Cornell University, 2015.
- [21] FELDMAN, J. B. and TOPAL0GLU, H., “Revenue management under the markov chain choice model,” *Working paper, School of ORIE, Cornell University*, 2014.
- [22] GALLEG0, G., IYENGAR, G., PHILLIPS, R., and DUBEY, A., “Managing flexible products on a network.” CORC Technical Report TR-2004-01, Department of Industrial Engineering and Operations Research, Columbia University, New York., 2004.
- [23] GALLEG0, G., RATLIFF, R., and SHEBALOV, S., “A general attraction model and sales-based linear program for network revenue management under customer choice,” *Operations Research*, vol. 63, no. 1, pp. 212–232, 2014.
- [24] GALLEG0, G. and TOPAL0GLU, H., “Constrained assortment optimization for the nested logit model,” *Management Science*, vol. 60, no. 10, pp. 2583–2601, 2014.

- [25] GALLEGO, G. and VAN RYZIN, G., “A multiproduct dynamic pricing problem and its applications to network yield management,” *Operations Research*, vol. 45, no. 1, pp. 24–41, 1997.
- [26] GALLEGO, G. and WANG, R., “Multiproduct price optimization and competition under the nested logit model with product-differentiated price sensitivities,” *Operations Research*, vol. 62, no. 2, pp. 450–461, 2014.
- [27] GRÖTSCHEL, M., LOVÁSZ, L., and SCHRIJVER, A., *Geometric algorithms and combinatorial optimization*. New York: Springer, 1988.
- [28] KUNNUMKAL, S. and TALLURI, K., “On the tractability of the piecewise-linear approximation for general discrete-choice network revenue management.” Working paper, School of IEOR, Columbia University, 2014.
- [29] KUNNUMKAL, S. and TALLURI, K., “Technical note a note on relaxations of the choice network revenue management dynamic program,” *Operations Research*, vol. 64, no. 1, pp. 158–166, 2016.
- [30] LEEMIS, L. M., “Nonparametric estimation of the cumulative intensity function for a nonhomogeneous poisson process,” *Management Science*, vol. 37, no. 7, pp. 886–900, 1991.
- [31] LEHMANN, E. L. and CASELLA, G., *Theory of Point Estimation*. New York: Springer, 2 ed., 1998.
- [32] LETHAM, B., LETHAM, L. M., and RUDIN, C., “Bayesian inference of arrival rate and substitution behavior from sales transaction data with stockouts,” *arXiv preprint arXiv:1502.04243*, 2015.
- [33] LI, G., RUSMEVICHIENTONG, P., and TOPALOGLU, H., “The d-level nested logit model: Assortment and price optimization problems,” *Operations Research*, vol. 63, no. 2, pp. 325–342, 2015.
- [34] LI, H. and HUH, W. T., “Pricing multiple products with the multinomial logit and nested logit models: Concavity and implications,” *Manufacturing & Service Operations Management*, vol. 13, no. 4, pp. 549–563, 2011.
- [35] LITTLEWOOD, K., “Forecasting and control of passenger bookings,” in *AG-IFORS 12th Annual Symposium Proceedings*, (Nathanya, Israel), pp. 95–117, 1972.
- [36] LIU, Q. and VAN RYZIN, G., “On the choice-based linear programming model for network revenue management,” *Manufacturing & Service Operations Management*, vol. 10, no. 2, pp. 288–310, 2008.
- [37] MÉNDEZ-DÍAZ, I., MIRANDA-BRONT, J. J., VULCANO, G., and ZABALA, P., “A branch-and-cut algorithm for the latent class logit assortment problem,” *Discrete Applied Mathematics*, vol. 164, pp. 246–263, 2014.

- [38] NEWMAN, J., FERGUSON, M., and GARROW, L., “Estimating discrete choice models with incomplete data,” *Transportation Research Record: Journal of the Transportation Research Board*, no. 2302, pp. 130–137, 2012.
- [39] NEWMAN, J. P., FERGUSON, M. E., and GARROW, L. A., “Estimating gev models with censored data,” *Transportation Research Part B: Methodological*, vol. 58, pp. 170–184, 2013.
- [40] NEWMAN, J. P., FERGUSON, M. E., GARROW, L. A., and JACOBS, T. L., “Estimation of choice-based models using sales data from a single firm,” *Manufacturing & Service Operations Management*, vol. 16, no. 2, pp. 184 – 197, 2014.
- [41] POWELL, W. B., *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. New York: John Wiley & Sons, 2007.
- [42] RATLIFF, R. M., VENKATESHWARA R., B., NARAYAN, C. P., and YELLEPEDDI, K., “A multi-flight recapture heuristic for estimating unconstrained demand from airline bookings,” *Journal of Revenue and Pricing Management*, vol. 7, no. 2, pp. 153–171, 2008.
- [43] RAYFIELD, W. Z., RUSMEVICHIENTONG, P., and TOPALOGLU, H., “Approximation methods for pricing problems under the nested logit model with price bounds,” *INFORMS Journal on Computing*, vol. 27, no. 2, pp. 335–357, 2015.
- [44] RUSMEVICHIENTONG, P., SHMOYS, D., TONG, C., and TOPALOGLU, H., “Assortment optimization under the multinomial logit model with random choice parameters,” *Production and Operations Management*, vol. 23, no. 11, pp. 2023–2039, 2014.
- [45] RUSMEVICHIENTONG, P., SHEN, Z.-J. M., and SHMOYS, D. B., “Dynamic assortment optimization with a multinomial logit choice model and capacity constraint,” *Operations research*, vol. 58, no. 6, pp. 1666–1680, 2010.
- [46] RUSMEVICHIENTONG, P. and TOPALOGLU, H., “Robust assortment optimization in revenue management under the multinomial logit choice model,” *Operations Research*, vol. 60, no. 4, pp. 865–882, 2012.
- [47] SHARKEY, T. C., ROMELJN, H. E., and GEUNES, J., “A class of nonlinear non-separable continuous knapsack and multiple-choice knapsack problems,” *Mathematical Programming*, vol. 126, no. 1, pp. 69–96, 2011.
- [48] SMITH, B. C., LEIMKUHLE, J. F., and DARROW, R. M., “Yield management at american airlines,” *Interfaces*, vol. 22, no. 1, pp. 8–31, 1992.
- [49] TALLURI, K., “A finite-population revenue management model and a risk-ratio procedure for the joint estimation of population size and parameters,” *Available at SSRN 1374853*, 2009.

- [50] TALLURI, K. and VAN RYZIN, G., “Revenue management under a general discrete choice model of consumer behavior,” *Management Science*, vol. 50, no. 1, pp. 15–33, 2004.
- [51] TALLURI, K. T. and VAN RYZIN, G. J., *The Theory and Practice of Revenue Management*. New York: Springer, 2004.
- [52] TAUCHEN, G., “Diagnostic testing and evaluation of maximum likelihood models,” *Journal of Econometrics*, vol. 30, no. 1, pp. 415–443, 1985.
- [53] TRAIN, K. E., *Discrete Choice Methods with Simulation*. Cambridge, UK: Cambridge University Press, 2009.
- [54] VAN RYZIN, G. and VULCANO, G., “A market discovery algorithm to estimate a general class of nonparametric choice models,” *Management Science*, vol. 61, no. 2, pp. 281–300, 2014.
- [55] VULCANO, G., VAN RYZIN, G. J., and CHAAR, W., “Choice-based revenue management: An empirical study of estimation and optimization,” *Manufacturing & Service Operations Management*, vol. 12, no. 3, pp. 371–392, 2010.
- [56] VULCANO, G., VAN RYZIN, G. J., and RATLIFF, R., “Estimating primary demand for substitutable products from sales transaction data,” *Operations Research*, vol. 60, no. 2, pp. 313–334, 2012.
- [57] WANG, R., “Assortment management under the generalized attraction model with a capacity constraint,” *Journal of Revenue & Pricing Management*, vol. 12, no. 3, pp. 254–270, 2013.
- [58] ZHANG, D., “An improved dynamic programming decomposition approach for network revenue management,” *Manufacturing & Service Operations Management*, vol. 13, no. 1, pp. 35–52, 2011.
- [59] ZHANG, D. and ADELMAN, D., “An approximate dynamic programming approach to network revenue management with customer choice,” *Transportation Science*, vol. 43, no. 3, pp. 381–394, 2009.
- [60] ZHANG, D. and COOPER, W. L., “Pricing substitutable flights in airline revenue management,” *European Journal of Operational Research*, vol. 197, no. 3, pp. 848–861, 2009.