Estimation and Optimization in Online Marketplaces

by

Hanwei Li

Submitted to the Institute of Data, Systems, and Society
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Social Engineering Systems

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2022

© Massachusetts Institute of Technology 2022. All rights reserved.

Author .....
Institute of Data, Systems, and Society
May 16 2022

Certified by .....
David Simchi-Levi
Professor of Engineering Systems, MIT
Thesis Supervisor

Certified by .....
Stephen Graves
Abraham J. Siegel Professor of Management, MIT
Thesis Committee Member

Certified by .....
Víctor Martínez de Albéniz
Professor of Production, Technology and Operations Management,
IESE Business School
Thesis Committee Member

Accepted by .....
Fotini Christia
Chair, Social and Engineering Systems Program

Estimation and Optimization in Online Marketplaces

by

Hanwei Li

Submitted to the Institute of Data, Systems, and Society
on May 16 2022, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Social Engineering Systems

Abstract

The emergence of e-commerce business models (such as Airbnb and Amazon) brings opportunities and challenges to their operations. This thesis studies several estimation and optimization problems within the online platform domain, using data-driven approaches in operations management.

The thesis consists of three components. Motivated by the unique setting of Airbnb, in the first work, we consider a game-theoretical setup in which each seller on the platform provides a single-unit product and competes with one another on price. We investigate sellers' optimal pricing decisions and the platform's optimal assortment display policy. We find that the platform should display the entire assortment to all the customers when demand is sufficiently high. Moreover, we propose a tabulation algorithm and a mixed-integer programming formulation to effectively solve for the sellers' and the platform's optimal decisions. Additionally, in the optimal display policy, we incorporate constraints to guarantee a certain degree of seller and customer fairness on both system and individual levels.

The second work is also closely related to marketplaces like Airbnb, where we estimate and optimize the impact of photo layout. We apply Resnet50, a convolutional neural network model, to build two separate, supervised learning models to evaluate the image quality and room types posted by Airbnb hosts. Then, we characterize the overall impacts of photo layout by the room type, photo quality, and display order. To address the estimation challenges in the Airbnb setting, we propose a novel pairwise comparison model to consistently estimate the impact of photo layout. Our estimation results suggest that the cover image has a significantly more significant impact than non-cover photos. A high-quality bedroom cover image leads to the most significant increase in demand. The counterfactual analysis shows the potential impact when adopting the optimal photo layouts.

In the third work, we collaborate with a global online fashion retailer, Zalando, to optimize large-scale price discount decisions. We address Zalando's local and global business challenges by applying a three-step process. We cluster products into groups that behave similarly and pre-solve the aggregated problem. In the second step, we decompose the problem using Lagrangian relaxation into a problem for each product

(SKU) and provide an efficient way to identify the Lagrange multipliers. Finally, we optimize decisions for individual products addressing local business constraints. For this new approach, which was implemented as part of Zalando's price discount decision process, we provide results from offline tests and field experiments to demonstrate its benefit.

Thesis Supervisor: David Simchi-Levi
Title: Professor of Engineering Systems, MIT

Thesis Committee Member: Stephen Graves
Title: Abraham J. Siegel Professor of Management, MIT

Thesis Committee Member: Víctor Martínez de Albéniz
Title: Professor of Production, Technology and Operations Management, IESE Business School

Acknowledgments

This thesis would have not been possible without the extraordinary people who have supported me over the years.

First and foremost, I would like to thank my advisor, David Simchi-Levi. It is my great honor to be advised by David. I still remember when we first talked over Skype, and first met in person at MIT. I have been admiring his sharpness, kindness, and passion since then. Over the years, I have learned so much from his experience, insights, and guidance in research. I am also amazed by his charisma as my role model beyond research, teaching, industry collaborations, and everyday life.

Next, I would like to thank my thesis committee members, Steven Graves and Víctor Martínez de Albéniz. I met Steve in my undergraduate school SUTD before joining MIT. He has always been supportive and insightful both within and beyond research. I share very similar research interests with Víctor and had read many of his works before we met in Barcelona when I visited IESE. He shares valuable suggestions in all of my thesis works, as well as my academic career.

I appreciate the support from our department IDSS. As the first batch of Ph.D. students, I would like to thank Munther Dahleh, Ali Jadbabaie, and all IDSS faculties for making IDSS and SES possible. I also appreciate the tremendous support from IDSS staff, especially Janet Kerrigan and Elizabeth Milnes.

This thesis would not have been possible without the collaboration of my great coauthors. I want to thank Weiming Zhu for opening the door of the empirical field and guiding me through it. We have been working closely in research and beyond research as friends. I have also received tremendous help from Michelle Wu in research and industry projects. She has also been a caring mentor in my life. I want to thank my lab mate coauthors, Rui Sun and Hongyu Chen, for their excellent thoughts, valuable discussions, and the time working together as friends. I also appreciate the funding support and my coauthors from AB Inbev and Zalando.

I also want to thank all my friends for always being supportive. Finally, I owe my deepest gratitude to my parents. This thesis is dedicated to them.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

The field of Operations Management focuses on the business operations decision-making process by increasing its efficiency. With technological advances, the emergence of new business models, like e-commerce markets, brings both opportunities and challenges to their operations. The widespread COVID-19 pandemic magnifies such challenges. Covid-19 adds $219 billion to US e-commerce sales in 2020-2021 (Berthene 2022). Koetsier (2020) estimates that COVID-19 Accelerated e-commerce growth for four to six years.

There are several distinctions between e-commerce platforms and traditional brick-and-mortar retailing stores. These distinctions impose various challenges on the e-commerce decision-making process. Firstly, online marketplaces may involve multiple stakeholders, including customers and third-party sellers. The decision-making process is much more complicated than a single decision maker's problem, as in the traditional retailing setting. The existence of multiple sellers creates competition on quality or price, which is generally challenging to characterize in reality. On top of this, the platform needs to carefully decide on the display policy for the sellers to optimize the expected revenue. As a result, the flexibility in the platform's display policy may lead to unfairness issues across sellers or customers, which also needs to be addressed.

Secondly, unlike brick-and-mortar stores where customers can access the products physically, e-commerce heavily relies on display and visual information. An example would be Airbnb, a peer-to-peer lodging marketplace provider. Customers browse online the features of the lodging options without physically accessing the option. As a result, the web page's information, including images and reviews from other customers, becomes highly significant. Unlike the traditional data format (like prices), images and texts contain much more information, imposing challenges in the estimation process.

The third difference is that e-commerce platforms are not constrained by limited shelf spaces, leading to a huge assortment or seller base. Pricing decisions are required for multiple markets and time periods on a huge assortment. Moreover, there usually exist business constraints to couple the pricing decisions across the products. As a result, products cannot be priced individually, and the joint optimization problem has a very large scale.

This thesis investigates estimation and optimization challenges in several e-commerce settings. An overview of the works in this thesis is as follows.

1.2 Overview

In Chapter 2, we study the assortment display problem in online marketplaces, in the existence of price competition and fairness concerns. Online platforms have been expanding the seller base to widen their product assortment to match consumers' individual preferences. Nevertheless, the increasing number of sellers leads to intensified competition and results in sellers setting lower prices for the products. Thus, it is unclear whether displaying all the sellers to the entire customer base maximizes platform revenue. Motivated by the unique setting of Airbnb, we consider a game-theoretical setup in which each seller on the platform provides a single-unit product and competes with one another on price. We investigate sellers' optimal pricing decisions and the platform's optimal assortment display policy, which is characterized by the partitioning of products and traffic assigned to each partition. We find that the

platform should display the entire assortment to all the customers when demand is sufficiently high. Moreover, we propose a tabulation algorithm and a mixed-integer programming formulation to effectively solve for the sellers' and the platform's optimal decisions. Additionally, we incorporate constraints to guarantee a certain degree of the seller and customer fairness, on both system and individual levels, in the optimal display policy. Using data from Airbnb, we present a case study to illustrate how our model framework can be applied in practice. Finally, we extend the case in which each seller supplies a distinct product with an inventory size of one by considering scenarios in which each product has more than one unit.

The second work in Chapter 3 is also closely related to marketplaces like Airbnb, where we estimate and optimize the impact of photo layout. Host-generated property images as a visual channel reveal important information about properties. Selecting proper images to display can lead to higher demand and increased rental revenue. In this paper, we define, estimate, and optimize the impacts of Airbnb photos on customers' renting decisions. We apply Resnet50, a convolutional neural network model, to build two separate, supervised learning models to evaluate the image quality and room types posted by Airbnb hosts. Then, we characterize the overall impacts of photo layout by the room type featured in the photo, photo quality, and the order of display on the listings' webpages. To address two estimation challenges in the Airbnb setting, namely censored demand and changing consideration sets, we propose a novel pairwise comparison model that utilizes customers' booking sequence data to consistently estimate the impact of photo layout on customers' renting decisions. Our estimation results suggest that the cover image has a significantly larger impact than non-cover photos and that a high-quality bedroom cover image leads to the most significant increase in demand. Furthermore, we build a non-linear integer programming optimization problem and develop an algorithm to determine the optimal photo layout. Our counterfactual analysis suggests that a listing's unilateral adoption of optimal photo layout leads to 11.0% more bookings on average. Moreover, depending on the neighborhood and market size, when listings simultaneously switch to the optimal photo layout, they get booked for two to five additional days

in a year on average, which boosts the revenue by $500 to $1100.

In the third work in Chapter 4, we collaborate with a global online fashion retailer, Zalando, as an example of how a global retailer can utilize a massive amount of data to optimize price discount decisions over a large number of products in multiple countries on a weekly basis. Given demand forecasts under a collection of discrete prices, Zalando's objective is to set discount levels to maximize total profit over the entire selling horizon while taking into account both *local* and *global* business constraints. Local constraints refer to single product level requirements, where Zalando needs to balance sales across different countries and over different weeks while adhering to a first-come-first-serve policy. As long as product inventory exists, a customer is served independent of the customer's origin country or time of arrival. Global constraints refer to specific targets set by management for different product categories and each country. We address these challenges by applying a three-step process. In the first step, we cluster products into groups that behave similarly and solve the aggregated problem to allow us to decouple the problem into a problem for each product category. Each product category includes thousands of individual products (SKUs) and the various markets where products are sold, each with its own target sales and margins. In the second step, we decompose this problem using Lagrangian relaxation into a problem for each product (SKU) and provide an efficient way to identify the Lagrange multipliers. Finally, in the last step, we optimize pricing decisions for individual products and also address local business constraints. For this new approach, which was implemented as part of Zalando's price discount decision process, we provide results from offline tests and field experiments to demonstrate its benefit.

Finally, Chapter 5 summarizes the contributions and discusses the potential future directions. The technical proofs and supplemental material for each chapter are included in the appendices.

Chapter 2

Assortment Display, Price Competition, and Fairness in Online Marketplaces

2.1 Introduction

It has become a norm for online platforms to widen their seller base constantly. For example, Airbnb now offers 5.6 million listings worldwide[1]; Amazon has 6.3 million total sellers worldwide, 1.5 million of whom are active[2]. Such rapid expansion of sellers is partly due to the network nature of platforms. Yet, it also reflects the platforms' effort to satisfy the heterogeneous tastes of their customers with a constellation of products. Unlike brick-and-mortar retailers whose product assortment is limited by shelf space, online platforms can increase the number of sellers and the size of the product assortment with little cost. This enables the large platforms to further expand their assortments.

Although a large seller base increases the likelihood the consumer will find the preferred variety, it may also jeopardize platform revenue because sellers would set lower prices for their products in response to the resulting heightened competition. Real-

[1]https://news.airbnb.com/about-us/

[2]https://www.marketplacepulse.com/amazon/number-of-sellers

izing that displaying all the products to the entire customer base may not maximize platform revenue, some companies start to ramp up sales by prioritizing a subgroup of products on the search results pages, which spurs concerns over the unfairness of the product display.[3] In this paper, we focus on a unique setting in which each seller provides a different product with one-unit inventory per time period (e.g., Airbnb, eBay), we contribute to this growing discussion in the literature on assortment optimization by answering the following questions: (i) how does competition affect the seller's pricing decision when each seller supplies one unit of a different product, (ii) when would showing different subsets of products to different subsets of customers generate higher platform revenue than that generated by the current practice and (iii) how to design an assortment display policy that is fair to the sellers and customers, and what is the associated cost of implementing such a policy.

To answer the above questions, we construct a game-theoretical model. The platform decides whether and how each seller should be grouped into different partitions and how much traffic each partition should receive to maximize the total revenue. Furthermore, each seller supplies a different product with one-unit inventory and sets the product's price in response to other sellers' prices within the same partition. Under this setup, we theoretically demonstrate that it is optimal to display the entire product assortment when a platform faces sufficient demand. When the platform demand is moderate, however, the derivation of the equilibrium price requires solving a system of nonlinear equations, which can be computationally infeasible because the number of display policies increases rapidly with the number of sellers. To derive the equilibrium price for each seller and ultimately prescribe the display policy, we propose a novel approach to calculate the equilibrium price under flexible display constraints. Specifically, we first discretize and tabulate the left- and right-hand sides of the first-order conditions (FOCs) of the seller's pricing decision. Incorporating the outputs from the tabulation procedure, we rewrite the platform's problem as a mixed-integer programming (MIP) model. The MIP framework can effectively solve the optimal display policy and can easily be adjusted to include new constraints that

[3]https://www.nytimes.com/2018/06/23/business/amazon-the-brand-buster.html

Table 2.1: Summary of Optimal Display Policies Under Different Settings and Demand Levels

	Demand/Supply Ratio		
	Very High	Moderate	Very Low
Unit Inventory	Display Everything	MIP†	Multiple Partitions
Limited Inventory	Display Everything	MIP	Multiple Partitions
Infinite Inventory	Display Subset	Display Subset	Display Single Item

† MIP indicates that the specific display policy is inconclusive and needs to be advised by the solution of the MIP problem. The optimal display strategies under the remaining market conditions are known, yet the specific partitions of sellers and customers still require solving the MIP problem.

reflect restrictions on the display policy.

Leveraging our game-theoretical model and the MIP framework, we summarize the optimal display policy under different market conditions in the first row of Table 2.1. Similar to Heese and Martínez-de Albéniz (2018), we note that it is sometimes economically sensible to display only a subset of products to the entire customer base. While such a policy might be practical in offline retail settings, in which the limited shelf space naturally prevents the store from displaying all the products, this display policy essentially denies a proportion of the sellers the ability to join the platform and could spur concerns over an unfair business environment on online marketplaces. To guarantee a certain degree of seller fairness, we introduce the notion (α, δ)-fairness on the system (partition) level and incorporate such fairness constraints into the MIP framework. Here, α is defined as the closeness in the attractiveness of each partition and δ the closeness in the traffic allocated to each partition. Moreover, we define the 'envy level' as a seller's revenue difference when moving to other partitions as a fairness measure on the individual seller level. We showcase how the upper bound on a seller's envy level is determined by the combination of the (α, δ). Meanwhile, to ensure that customers perceive the partitions as fair, we introduce the concept of γ-fairness to measure the relative difference in customer welfare across partitions. Similar to (α, δ)-fairness, we also showcase that γ-fairness can be incorporated into our MIP framework.

Additionally, we present a case study that draws transaction data from Airbnb

to showcase how our framework can be applied in practice. Specifically, we estimate the quality of each listing by minimizing the squared discrepancy between each listing's rental probability and the realized demand outcome. We then conduct a counterfactual analysis under different demand scenarios to recommend the optimal display policy for the platform. Furthermore, we calculate the revenue loss when the platform shifts from the $(0, 0)$-fair display, i.e., the unrestricted case, to $(0, 1)$-fair, in which each partition receives equal demand. The platform incurs less than a 20% loss in revenue when the demand level is low or moderate. Nevertheless, the revenue gap ceases when the demand level is high, as the optimal policy under the unrestricted case is to display all the sellers to all customers, which is also $(0, 1)$-fair.

Finally, we extend the case in which each seller supplies a distinct product with the inventory of size one by considering two scenarios where the product each seller provides has (i) limited inventory or (ii) infinite inventory. We formulate the objective function for each seller and demonstrate that when facing sufficient demand, it is still optimal to display all the products when the seller possesses multiple units of the product, but the platform should display only the highest-quality item when each seller has infinite units of each product, and the number of sellers is large. The insights for optimal display policies under different settings are summarized in Table 2.1.

2.2 Literature Review

This work is related to three streams of literature, namely, assortment optimization, price competition, and fair resource allocation.

The field of assortment optimization has been active in recent decades, including both dynamic assortment planning and static assortment display. Static assortment display only focuses on the optimal assortment policy for a single period, and Kök et al. (2008) provides a comprehensive review of the literature in this field. Talluri and Van Ryzin (2004) show that under the MNL demand model, the optimal assortment set is that with the highest profit margins. However, this structural property no

longer holds in our setting, as the limited inventory size could lead to unsatisfied customers, which a larger assortment could alleviate. Under the static setup, the interplay between assortment and pricing has also been studied extensively. Anderson et al. (1992) show that it is optimal for the retailer to use the same markup for all products, although this result no longer holds when price sensitivities differ across products (e.g., (Wang 2012) or (Gallego and Wang 2014)). In this paper, we model a similar process as a two-stage game in which the platform decides the assortment strategy. Dynamic assortment optimization focuses on optimal product planning over time under various customer behaviors; see Ferreira and Goh (2021), Bernstein et al. (2015), Caro et al. (2014), Bernstein and Martínez-de Albéniz (2017), Lei et al. (2021), and Wagner and Martínez-de Albéniz (2020).

In the traditional assortment planning literature, prices are usually determined by a centralized decision-maker. By contrast, this paper models the price competition among sellers in online marketplaces. Price competitions are prevalent and are often combined with assortment problems (e.g., (Caro and Martínez-de Albéniz 2012), (Federgruen and Hu 2015)). (Besbes and Sauré 2016) consider the game-theoretic setup where two retailers compete on assortment and pricing and show that competition would lead to a larger assortment breadth. Meanwhile, the closest work to our paper is Heese and Martínez-de Albéniz (2018), in which the assortment planning process is modeled as a two-stage game between a retailer and its upstream manufacturers. They engage in horizontal price competition with each other. Different from Heese and Martínez-de Albéniz (2018) which mainly analyzes the optimal fixed subset display policies, our work focuses on designing the optimal partition display policies. Moreover, our sellers in each partition compete through price to maximize their revenue, while in Heese and Martínez-de Albéniz (2018), the upstream manufacturers compete for a fixed amount of slots in the assortment.

Our paper discusses the optimal display policy under several fairness constraints and is thus related to previous literature investigating the tradeoff between fairness and efficiency. Most of the previous work that explores the optimal decisions under several fairness constraints adopts the notion of α-fairness proposed in Mo and

Walrand (2000). Specifically, Mo and Walrand (2000) provides a generalized definition of fairness using a parametric function, for which higher values of α indicate a fairer allocation, and the parameters $\alpha = 0, 1$, and ∞ correspond to the utilitarian, proportional and max-min fairness, respectively. Using the definition of α-fairness, Bertsimas et al. (2011, 2012) study the efficiency loss incurred to achieve a certain fairness level in offline resource allocation problems; McCoy and Lee (2014) quantify how equity and efficiency interact for humanitarian and health delivery supply chains, and Bateni et al. (2018) investigate an online resource allocation problem on a marketplace and provide empirical validation of the proposed allocation algorithm. Our work also uses the notion of α-fairness, but our notion of α-fairness departs from the traditional definition. It requires instead the relative ratio of incoming traffic and the total competitiveness between any two partitions to remain within a specific range in a way similar to Cohen et al. (2019).

2.3 Model Framework

2.3.1 Price Competition under Full Display

We first consider a simple setup where a platform (e.g., Airbnb) hosts N sellers (e.g., listing owners) and displays everyone by default. Each seller i supplies a single-unit product with quality a_i on the platform and observes the qualities of all the other competitors $\boldsymbol{a_{-i}}$. We denote the expected number of customer arrivals by M, which we assume is common knowledge to the platform and the sellers. Observing $(a_i, \boldsymbol{a_{-i}})$ and M, every seller i simultaneously decides the price for its product to maximize revenue. The pricing game is characterized as follows. We assume that M customers arrive simultaneously and observe the full product assortment. The utility that customer m obtains from product i is $u_{im} = a_i - \beta p_i + \epsilon_{im}$, where a_i is the quality of the product, β is the coefficient for price p_i and ϵ_{im} follows a $Gumbel(0, 1)$ distribution that captures customers' heterogeneous tastes. This way, customers select products according to the multinomial logit (MNL) choice model. Notably, as each product

has only unit inventory, if a product is requested by multiple customers, the seller will accept the offer from only one buyer, and the remaining customers will be rejected and unable to select another product. In this setting, seller i's payoff function can be written as:

$$\max_{p_i} \Pi_i(M, \boldsymbol{p_{-i}}) = p_i \left(1 - \left(1 - \frac{\exp\left(a_i - \beta p_i\right)}{1 + \sum_{j=1}^{N} \exp\left(a_j - \beta p_j\right)} \right)^M \right) . \tag{2.1}$$

Equation (2.1) states that seller i's expected revenue is the product of the price and the probability that product i is selected by at least one customer, which equals one minus the probability that all M customers choose options other than i. The utility of the outside option is normalized to one as in the classic MNL setup.

For brevity, we first denote the attractiveness of product i as $v_i = \exp\left(a_i - \beta p_i\right)$ and write the probability that a customer chooses product i as $q_i = v_i/(1 + \sum_{j=1}^{N} v_j)$. In this way, the price equilibrium, $p_i^*, i = 1, ..., N$, can be characterized by the FOC of Equation (2.1) as:

$$(1 + M\beta p_i^* q_i^*)(1 - q_i^*)^M = 1, \quad \forall i = 1, ..., N. \tag{2.2}$$

We next describe the properties of the equilibrium prices. To start, the pricing game always permits a unique equilibrium, which we formally state in the following Proposition 2.1:

Proposition 2.1. *A unique pure-strategy Nash equilibrium always exists for the seller's pricing game.*

This property guarantees that there always exists one and only one equilibrium regardless of the number of sellers and customers. In addition to existence and uniqueness, the equilibrium price has the following characteristics:

Proposition 2.2. *The equilibrium price p_i^* is increasing in a_i, i.e., products with higher qualities will charge higher prices. Furthermore, $\lim_{M\to\infty} \partial p_i^*/\partial a_i = 1/\beta$, $\forall i = 1, ..., N$.*

Proposition 2.2 has two implications. First, sellers with higher product quality charge higher prices in equilibrium. Second, the price change is linear in quality in the limiting case in which the platform faces considerable demand. As a result, all the products share the same attractiveness $a_i - \beta p_i^*$. While Propositions 2.1 and 2.2 characterize the properties of the equilibrium price, they do not reveal what equilibrium price each seller should set. We showcase how the exact value of the equilibrium prices, which is the solution to a system of nonlinear FOCs, can be computed in Section 2.4.

2.3.2 Price Competition under Partitioned Display

Different from the standard practice in which the platform displays all the sellers, we now consider a case in which the platform has the freedom to choose which product to display and to which customers. Specifically, we consider a two-stage game with the platform as the leader and the sellers as the follower. In the first stage, the platform announces a display policy $(\mathcal{S}_k, M_k)$ to the sellers. The display policy has two components. First, it describes how the platform partitions products from the full assortment into different subsets $\mathcal{S}_k$. Second, it indicates the number of customers M_k assigned to each product partition $\mathcal{S}_k$[4]. Importantly, in the rest of the paper, we refer to a display policy as the division of the assortment $\mathcal{S}$ and demand M that satisfies the following two conditions:

1. Each seller is assigned to one and only one partition, i.e., $\cup_k \mathcal{S}_k = \mathcal{S}$ and $\mathcal{S}_i \cap \mathcal{S}_j = \emptyset$.

2. Every customer can observe one and only one partition, i.e., $\sum_k M_k = M$.

Intuitively, there should not be overlaps of seller or customer groups. Allowing sellers to be assigned into multiple partitions would lead to multiple optimal prices for the same product. We assume each seller belongs to a single partition to exclude the potential of such price discrimination. This class of display policy is pretty general.

[4]In reality, as customer arrivals can be perceived as uniform within a short time window (e.g., every 10 seconds), the traffic allocation can be achieved by common modulo operations.

For instance, subset display is a special case of partition display that allocates all the customers to a single partition.

In the second stage of the game, after observing the display policy and $(a_i, \boldsymbol{a_{-i}})$, the qualities of all other competitors with whom seller i displays together in the same partition, every seller i maximizes its expected utility by simultaneously setting p_i. Finally, customers assigned to each partition arrive and make purchase decisions after observing the attractiveness of each product in that partition.

We now elaborate on seller i's optimization problem facing the partitioned display. In the second stage, each seller i in each assortment $\mathcal{S}_k$ observes the quality of the competing sellers within the same partition. After observing the quality for each seller $(a_i, \boldsymbol{a_{-i}}), i \in \mathcal{S}_k$, total arrival M_k traffic of the subset, and a belief on competitor prices p_{-i}, seller i faces the following optimization problem:

$$\max_{p_i} \Pi_i^s(\mathcal{S}_k, M_k, p_{-i}) = p_i \left(1 - \left(1 - \frac{\exp(a_i - \beta p_i)}{1 + \sum_{\{j \in \mathcal{S}_k\}} \exp(a_j - \beta p_j)} \right)^{M_k} \right). \tag{2.3}$$

Similar to the scenario of displaying everything, we can write the FOC for Equation (2.3) as:

$$(1 + M_k \beta p_i^* q_i^*)(1 - q_i^*)^{M_k} = 1, \quad \forall i \in \mathcal{S}_k. \tag{2.4}$$

Notably, both Propositions 2.1 and 2.2 hold valid for Equation (2.4), as each seller's equilibrium price is determined only by the competing sellers in the same partition, not by the equilibrium prices in other partitions. Foreseeing the equilibrium prices set by each seller, the platform decides the optimal partition to maximize the total revenue. Typically, the primary revenue source of platforms such as Airbnb or Uber is through commission, i.e., the platform takes a fixed percentage cut of the total revenue from each seller. Thus, omitting the fixed commission rate, the platform's

optimization problem can be formulated as

$$\begin{aligned}
&\max_{\mathcal{S}_k, M_k} \sum_k \sum_{i \in \mathcal{S}_k} \Pi_i^*(\mathcal{S}_k, M_k) \\
&\quad s.t. \cup_k \mathcal{S}_k = \mathcal{S}, \quad \mathcal{S}_i \cap \mathcal{S}_j = \emptyset \\
&\qquad \sum_k M_k = M \\
&\qquad \Pi_i^*(\mathcal{S}_k, M_k) = p_i^*(1 - (1 - q_i^*)^{M_k}) \\
&\qquad (1 + M_k \beta p_i^* q_i^*)(1 - q_i^*)^{M_k} = 1, \quad \forall i \in \mathcal{S}_k.
\end{aligned} \tag{2.5}$$

The first two constraints in Equation (2.5) manifest our assumptions that each product exclusively belongs to one subset and that each customer observes only one such subset of products. The last two constraints capture the incentive compatibility (IC) of the seller's pricing decision and imply that each seller sets the optimal price given a set of display policies $(\mathcal{S}_k, M_k)$. Intuitively, partitioning the entire assortment into several distinct product sets dampens the competition in the whole market, thus inducing sellers to charge higher prices. However, because the demand allocated to each partition is also lower than that under the full display policy, the probability that each product is selected by a customer can be lower. Thus, the overall impact of product partitioning on equilibrium pricing and the resulting platform revenue is unclear. In what follows, we provide a theoretical analysis of the optimal display strategy when the total demand M is sufficiently large, which is often the case for large online marketplaces such as Airbnb. First, we can characterize the following limiting behaviors:

Lemma 2.1. *For a fixed partition display $\mathcal{S}_k$ and any product $i \in \mathcal{S}_k$, define $p_i^*(\mathcal{S}_k, M_k)$ and $q_i^*(\mathcal{S}_k, M_k)$ to be the equilibrium price and selection probability of product i under partition $\mathcal{S}_k$ and demand M_k. Then, in the limiting case in which $M_k \to \infty$, we have*

1. $\lim_{M_k \to \infty} p_i^*(\mathcal{S}_k, M_k) = +\infty, \lim_{M_k \to \infty} q_i^*(\mathcal{S}_k, M_k) = 0$
2. $\lim_{M_k \to \infty} \frac{M_k q_i^*(\mathcal{S}_k, M_k)}{\ln(-\ln q_i^*(\mathcal{S}_k, M_k))} = 1.$

Lemma 2.1 describes the order of the equilibrium purchasing probability q_i^* in

the limiting case, which we use to calculate the revenue a product generates under different partitions as stated in Lemma 2.2.

Lemma 2.2. *Consider two subsets, $\mathcal{S}_1$ and $\mathcal{S}_2$, and a product i that is in both subsets, i.e., $i \in \mathcal{S}_1 \cap \mathcal{S}_2$. If the number of customers assigned to each subset maintains a fixed ratio $\gamma > 0$, then $\lim_{M\to\infty} \Pi_i^*(\mathcal{S}_1, M) - \Pi_i^*(\mathcal{S}_2, M/\gamma) = \ln \gamma$.*

Note that $\ln \gamma$ is strictly positive as long as $\gamma > 1$. Thus, Lemma 2.2 implies that in the limiting case, as long as the number of customers assigned to partition i remains a fixed proportion of the total number of customers, presenting product i to the entire customer base generates higher revenue than that to a fraction of the incoming traffic. This result holds regardless of how many other products are also included in subsets S_1 and S_2. Nevertheless, it is essential to note that Lemma 2.2 does not imply that the platform should not partition over sellers and customers. This is because Lemma 2.2 does not shed light on the optimal size the platform should set for each subset S_k, and the number of customers M_k rationed to each subset of products may not always remain at a constant ratio of the total incoming traffic. To complement Lemma 2.2, we provide the following result about the monotonicity of seller revenue on M.

Lemma 2.3. *For a fixed assortment display $\mathcal{S}$, revenue $\Pi_i(\mathcal{S}, M)$ is monotonically increasing in M.*

Combining Lemmas 2.2 and 2.3, we now provide the sufficient condition to make it optimal for the platform to display everything as a whole when demand becomes high through Theorem 2.1.

Theorem 2.1. *Suppose that there are N products $\mathcal{S} = \{1, 2, \ldots, N\}$. For any $\gamma > 1$, there exists a threshold $M(\gamma)$, such that when $M > M(\gamma)$, for any display policy $\{\mathcal{S}_k, M_k\}_{k=1}^K$ that satisfies $M_k < M/\gamma$, we have*

$$\sum_{i=1}^{N} \Pi_i^*(\mathcal{S}, M) > \sum_{k=1}^{K} \sum_{i\in\mathcal{S}_k} \Pi_i^*(\mathcal{S}_k, M_k).$$

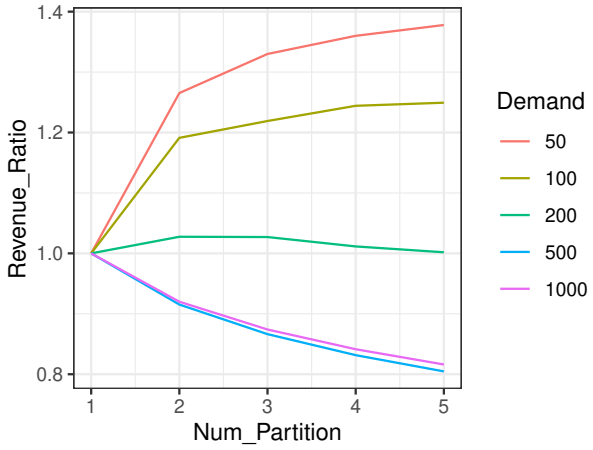

Figure 2-1: Revenue under Partition Display with Equal Demand

Proof for Theorem 2.1. As the number of subsets $\mathcal{S}_1 \subset \mathcal{S}$ is finite, according to Lemma 2.2, when $M > M(\gamma)$, where $M(\gamma)$ is the threshold that depends on γ, we have

$$\Pi_i^*(\mathcal{S}, M) - \Pi_i^*(\mathcal{S}_1, M/\gamma) > 0, \quad \forall \mathcal{S}_1 \subset \mathcal{S}, i \in \mathcal{S}_1.$$

Lemma 2.3 further states that $\Pi_i(\mathcal{S}_k, M)$ is monotonically increasing in M, and we can conclude that

$$\begin{aligned}\sum_{i=1}^{N} \Pi_i^*(\mathcal{S}, M) &\geq \sum_{k=1}^{K} \sum_{i \in \mathcal{S}_k} \Pi_i^*(\mathcal{S}_k, M/\gamma) \\ &\geq \sum_{k=1}^{K} \sum_{i \in \mathcal{S}_k} \Pi_i^*(\mathcal{S}_k, M_k).\end{aligned}$$

It is helpful to discuss the implication behind Theorem 2.1. For a platform facing sufficiently high demand, the platform enjoys the highest total revenue when the platform exhibits all products concurrently to all customers, i.e., there is no need to group products into different partitions and display each partition to a subset of customers. Importantly, for large platforms on which most sellers offer a single-unit product, Theorem 2.1 provides a theoretical justification that displaying the entire assortment to all customers leads to the highest revenue. To verify Theorem 2.1 and to gain insight into the optimal display policy under an arbitrary M, we perform numerical analysis and present the results in Figure 2-1. [5] Specifically, Figure 2-1

[5]The detailed derivation of equilibrium prices and optimal display policies will be introduced in

compares the optimal total revenue under different partition scenarios. We select an instance with $N = 100$ sellers and let each partition receive the same level of demand M/K. $K = 1$ corresponds to the default policy that displays all sellers to all customers. We normalize the revenue of the default policy to one and compare the revenue under different display policies under various demand levels. When the demand level is low ($M = 50$ and $M = 100$), dividing the sellers into more partitions is more profitable. When the demand level is high ($M = 500$ or $M = 1000$), it is optimal to maintain the display-everything policy, which is consistent with the conclusion of Theorem 2.1.

2.3.3 Discussion on Model Assumptions

We have introduced several assumptions in modeling the decision-making process in online marketplaces. We now discuss these assumptions before we analyze the optimal display policy.

Customer's Simultaneous Arrival and Multiple Bookings In our model, we assume that customers arrive simultaneously. While customers arrive sequentially over a time span, the purchasing or booking requests may not be satisfied immediately on platforms such as eBay or Airbnb. For example, many Airbnb hosts would turn off the 'instant booking' option so that they get to evaluate and screen the potential tenants. Usually, hosts would review the booking requests regularly, say at the end of every day, and eventually accept one request. We model the unsatisfied demand in case of multiple requests to be lost, and this is supported by Fradkin (2017), which suggests that 42% of customers' inquiries are rejected by hosts, and those rejected are 43% to 70% less likely to submit another request on the platform.

Product Filter When facing a large assortment, customers usually use filters to narrow search results and reduce search costs. After filtering, the displayed products are similar in one or several attributes and can thus be considered substitutes that engage in price competition with each other. As all products (sellers) compete through price in our model, these products can be viewed as the search results after filtering.

Section 2.4.

Case in point, we apply our framework to the Airbnb data in Section 2.6, in which we use listings within the same region and with the same room type as substitutes.

Product Visibility In our model, we assume products displayed in the same partition share the same visibility to the customers. That is, our formulation of product utility does not incorporate the specific page rank, i.e., the display order, of the product. As the page rank of each product may vary depending on the customer characteristics, we follow previous literature on assortment optimization in marketplaces (Dzyabura and Jagabathula 2018, Aouad et al. 2019), and do not model page rank for tractability of the model. Nevertheless, suppose the page rank of each product is invariant across customers and observed in the data. In that case, we can integrate the page rank information into each product's utility function to reflect its impact.

2.4 Derivation of the Optimal Display Policy

2.4.1 Characterizing the Price Equilibrium

In this section, we derive the platform's optimal partition strategy. To this end, we first characterize sellers' optimal pricing decisions under a given partitioning. Given an arbitrary display policy, the derivation of the equilibrium prices requires jointly solving a system of FOCs specified in Equations (2.4). These FOCs are highly nonlinear and cannot be characterized by the Lambert-W function, which Heese and Martínez-de Albéniz (2018) used to analyze the infinite inventory scenario. It is thus impossible to derive a closed-form solution for the equilibrium price p_i^*.

There are several viable approaches to solve for the equilibrium prices. One option is to compute through simulations. Since we have demonstrated the existence and uniqueness of the equilibrium, we can initialize a set of feasible prices and iteratively update each seller's price as the best response to the competitor's prices until they converge to an equilibrium where no seller has the incentive to deviate. However, this approach will no longer be applicable when the platform attempts to find the

optimal display policy. This is because the possible number of display policies grows exponentially with the number of sellers N, and enumerating all the policies becomes computationally burdensome, if not infeasible.

In our work, we propose a novel approach that transforms Equation (2.5) into an MIP problem to determine the induced optimal price and the optimal display policy. The idea is to define and discretize an intermediate $\mathcal{Z}$ in Equation (2.2) such that each FOC is decoupled from the system of FOCs given $\mathcal{Z}$. Then, we precalculate and tabulate the required inputs into tables and let the MIP solve for the display policy and the corresponding equilibrium prices that maximize the platform's total revenue. This approach enhances computational tractability by discretizing the nonlinear FOCs into a set of precomputed values. Moreover, the MIP framework provides a flexible infrastructure for the platform to impose different constraints on display policies. Granted, solving a large-scale MIP can still be computationally intensive, but we demonstrate in Appendix A.2 that our approach can be solved within a reasonable running time.

To better introduce the tabulation process, we first define the total attractiveness of all the sellers in partition k as $\mathcal{Z}^k = \sum_{i \in \mathcal{S}_k} \exp(a_i - \beta p_i)$. Furthermore, as each partition's equilibrium can be solved independently, we focus on solving the equilibrium for a single partition k and omit the superscript for ease of notation in the remainder of the section. In fact, there is a *unique* value of $\mathcal{Z}$ that can allow for a set of prices, $\{p_i\}$, to simultaneously satisfy Equation (2.4) and the definition of $\mathcal{Z}$. Moreover, under such $\mathcal{Z}$, the resulting set of prices represents the equilibrium price. Thus, we construct a feasible range $\mathcal{Z} \in [\underline{\mathcal{Z}}, \bar{\mathcal{Z}}]$, over which we discretize $\mathcal{Z}$ into L levels to form a ladder. For each seller $i \in N$ and each fixed value $\mathcal{Z}_j, j \in L$ in the ladder, we can compute the best response $p_{i,j}$ by applying common built-in root finding package to the FOCs. Although the FOCs are nonlinear, finding each root under a fixed $\mathcal{Z}_j$ reduces to a one-dimensional search problem that can be efficiently solved. By doing so, we can tabulate the best responses into a $N \times L$ table with each entry (i, j) corresponding to the equilibrium price for the i-th seller under the j-th value in the ladder of $\mathcal{Z}$. Similarly, we also track seller i's attractiveness

Algorithm 1: Tabulation of Formulation Input

input : $\{a_i,\ i = 1, ..., N\}$,$\{\mathcal{Z}_j,\ j = 1, ..., L\}$, M

for $j \in \{1, ..., L\}$ **do** // Iterate over $\mathcal{Z}$

  for $i \in \{1, ..., N\}$ **do** // Iterate over sellers

    Solve FOC

$$p^*_{i,j} = \arg_p \ \{(M\beta pq + 1)(1 - q)^M = 1\},$$

    to get the best response price $p^*_{i,j}$

    Tabulate $E_{i,j} = \exp(a_i - \beta p^*_{i,j})$

    Tabulate $\Pi_{i,j} = p^*_{i,j}\big(1 - (1 - \exp(a_i - \beta p^*_{i,j})/(1 + \mathcal{Z}_j))^M\big)$

  end for

end for

output: $\{p^*_{i,j}, E_{i,j}, \Pi_{i,j},\ i = 1, ..., N, j = 1, ..., L\}$

$E_{i,j} = \exp(a_i - \beta p^*_{i,j})$ and the induced revenue $\Pi_{i,j}$ in separate tables. These three tables $p^*_{i,j}, E_{i,j}$ and $\Pi_{i,j}$ $(i = 1, ..., N, j = 1, ..., L)$ are connected in the sense that the price in the (i, j)-th entry in the $E_{i,j}$ and $\Pi_{i,j}$ tables is identical to the value of the (i, j)-th entry in the table for $p^*_{i,j}$. We summarize the tabulation procedure in Table 1. This process allows us to formulate the problem as a mixed integer linear programming (MILP) that we introduce in the next subsection.

2.4.2 MIP Formulation and Fixed Subset Display

With the pretabulated inputs, we now construct the MILP model to solve the price equilibrium. We start with the simplest case, in which the platform displays all the products to all the customers as currently employed by platforms such as Airbnb and ebay. Given the qualities $\{a_i,\ i = 1, ..., N\}$, demand M, the total attractiveness $\{\mathcal{Z}_j,\ j = 1, ..., L\}$ and tabulated inputs $\{p^*_{i,j}, E_{i,j}, \Pi_{i,j},\ i = 1, ..., N, j = 1, ..., L\}$, the goal is to find a column j^* from table $E_{i,j}$, such that $\sum_i E_{i,j^*} = \mathcal{Z}_{j^*}$. The existence and uniqueness of the equilibrium are guaranteed by Proposition 2.1. For simplicity, we will use this exact form of constraints with equality in the remainder of the paper. However, in practice, it is unlikely that there exists a unique $\mathcal{Z}_{j^*}$ value that makes the equilibrium coincide with one point on the ladder. To ensure the MIP is feasible, we could replace the abovementioned equality constraint with two inequalities $\sum_i E_{i,k} \geq 0.5(\mathcal{Z}_k + \mathcal{Z}_{k-1})$ and $\sum_i E_{i,k} \leq 0.5(\mathcal{Z}_k + \mathcal{Z}_{k+1})$. Intuitively, the

solution will approach the true optimum as the discretization of the $\mathcal{Z}$ ladder becomes finer.

We compute the equilibrium prices under the default display policy in the example above. The platform does not need to make any decisions because the policy is fixed. We now present the case in which the platform chooses to display a fixed subset of the sellers. This scenario is a special case of our partition display policy allocating all the traffic M to a single partition. It is in line with the classic setting in the assortment optimization literature (e.g., Talluri and Van Ryzin 2004). Formally, we denote by z_j the binary decision variable on whether the j-th column is selected and by $x_{i,j}$ the binary decision variable indicating whether seller i is assigned to the group with z value z_j. The formulation is as follows.

$$\begin{aligned} \max_{x,z} \quad & \sum x_{i,j}\ \Pi_{i,j} \\ s.t. \quad & \sum_i x_{i,j} E_{i,j} = z_j \mathcal{Z}_j, && \forall j = 1, ..., L \\ & \sum_j z_j = 1, \\ & x, z \ \text{ binary} \end{aligned} \tag{2.6}$$

This formulation selects one column $z_j = 1$ in the table and sellers $x_{i,j} = 1$ within the column. The objective maximizes the total revenue of all the sellers, which aligns with the platform's incentive. The first constraint in Equation (2.6) ensures that the definition of $\mathcal{Z}$ is satisfied. In particular, $z_j = 0$ when column j is not selected. Thus, we must have $x_{i,j} = 0$ for all i, because the inputs $E_{i,j}$ are nonzero. When column j is selected, the model will select a subset of rows to satisfy the definition of $\mathcal{Z}$. Furthermore, the second constraint ensures that only one column will be selected, as only one fixed assortment will be displayed to the customers. Notably, similar to the full display case, the first constraint cannot be met exactly due to the discrete nature of $\mathcal{Z}$. To guarantee the feasibility of the MIP, in practice, we could relax the constraint by adding buffers of the following form: $\sum_i x_{i,j} E_{i,j} \geq 0.5 z_j (\mathcal{Z}_j + \mathcal{Z}_{j-1})$ and $\sum_i x_{i,j} E_{i,j} \leq 0.5 z_j (\mathcal{Z}_j + \mathcal{Z}_{j+1})$.

Exploiting Equation (2.6), we now conduct numerical analysis to explore the price equilibrium under the default display policy. We simulate a market with $N = 100$ sellers whose product quality is normally distributed and follows $a \sim \mathcal{N}(3, 1)$. We set the price coefficient to be $\beta = 1$ and vary the demand level from 50 to 1,000. Panel (a) in Figure 2-2 plots the relationship between the equilibrium price and product quality under different demand scenarios. Specifically, Panel (a) is consistent with Proposition 2.2 in that prices are monotonically increasing in quality, and when facing high demand, the relationship between price and quality becomes linear.

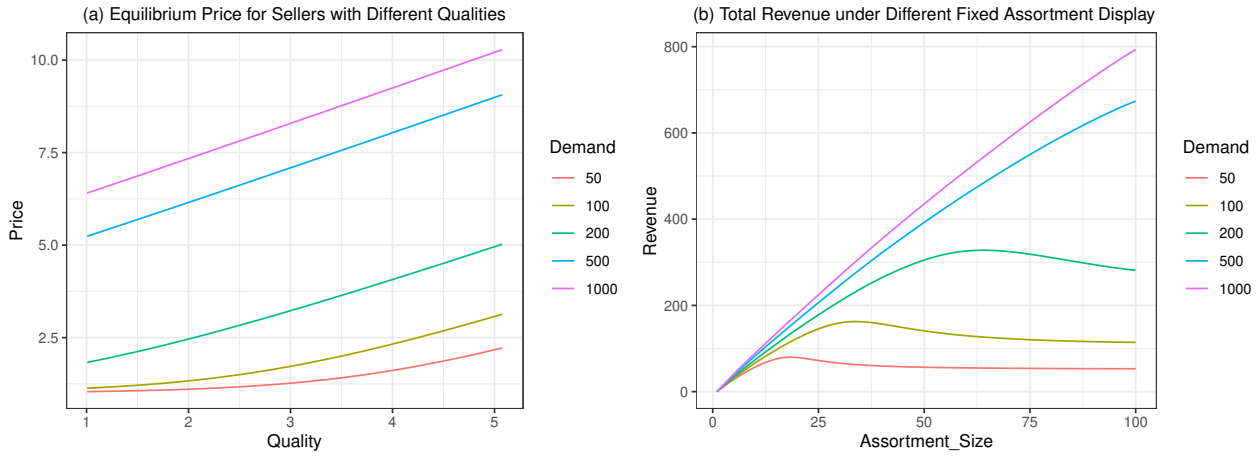

Figure 2-2: Price and Revenue under Fixed Assortment Display

We then conduct numerical analysis to study the total revenue under the optimal display policy given different assortment sizes. The results are presented in Panel (b) in Figure 2-2. Specifically, the x and y axes represent the cardinality of the displayed subset and the total revenue, respectively. The platform showcases the entire assortment to the customers at $x = 100$ and displays a subset over the remaining region. Interestingly, when demand is low, displaying a subset of products induces higher total revenue than displaying the whole assortment. As demand increases, the cardinality of the optimal assortment also grows. Consistent with the results from Theorem 2.1, it is optimal for the platform to display all N products when demand becomes sufficiently high. Notably, our numerical results provide important guidelines for platforms to decide on their expansion strategy. Newly launched platforms usually provide subsidies to incentivize service providers to join. However, our results suggest

that attracting a large number of service providers at an early stage can be suboptimal due to insufficient demand. Nevertheless, for large platforms that face heavy traffic on a daily basis, increasing the base of service providers can boost total revenue, as a reduction in the lost demand caused by (i) customers choosing the product and (ii) customers choosing the outside option outweigh the dampened product price caused by the more intense competition.

2.4.3 Partitioned Display

In Section 2.4.2, we proposed an MIP formulation to characterize the equilibrium price and the optimal display policy under fixed assortment. We now move to a more general case in which the platform has the flexibility to partition the products into arbitrary subsets and allocate the incoming traffic to each subset. To derive the optimal partition $(\mathcal{S}_k, M_k)$ that maximizes total revenue and characterizes the equilibrium price, we now describe the tabulation procedure and the MIP formulation based on Equation (2.5).

As the platform now has the flexibility to distribute the incoming demand unevenly to each product partition, each seller no longer faces M incoming customers when making its pricing decision. Thus, we first need to modify the tabulation procedure to incorporate flexibility in the number of customers each seller could potentially face. To this end, we normalize the total demand to one and discretize the fraction of demand (i.e., market share) from 0% to 100%. Specifically, we denote the v-th entry of the discretized market share by P_v. For example, when discretizing with 10% separation, we ultimately have 11 market share levels in total, and $P_1 = 0\%$ and $P_{11} = 100\%$. We then replace M with MP_v as the demand level and replicate the tabulation process described in Algorithm 1. In doing so, the outputs of the tabulation, $E_{i,j,v}$ and $\Pi_{i,j,v}$, will now have an additional dimension reflecting the MIP inputs under different fractions of market shares.

In addition to demand level v, the general class of partition display policy extends formulation (2.6) by allowing products to be grouped into different partitions, which we denote by k. The total partition number, K, is part of the inputs. In the main for-

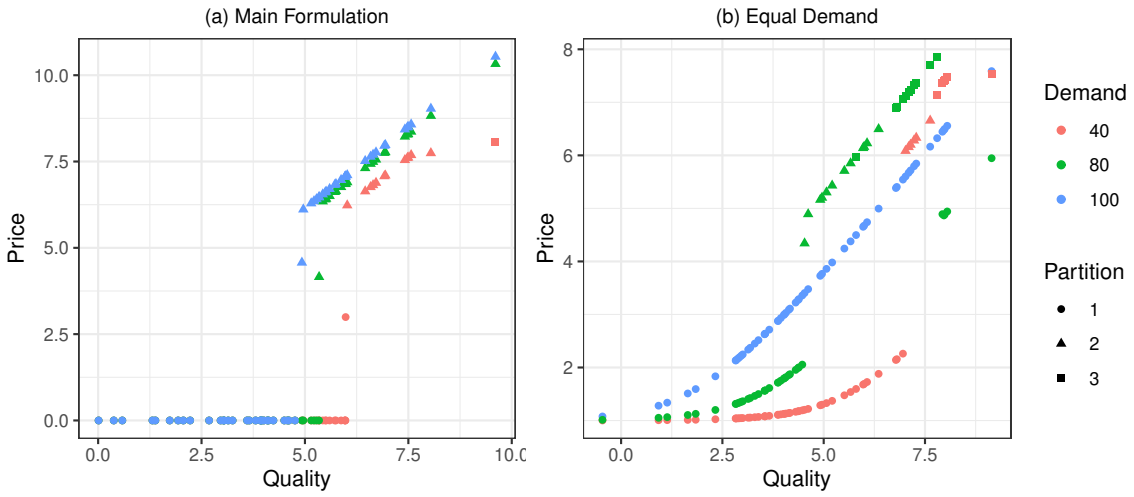

Figure 2-3: Price under different Levels of Demand M, at Optimal K

mulation, let $x_{i,j,k,v}$ denote the binary decision variable indicating whether to allocate seller i to partition k, which has total partition attractiveness $\mathcal{Z}_j$ and market demand MP_v. Similarly, $z_{j,k,v}$ denotes the binary decision variable of whether to select the jth level of $\mathcal{Z}$ and vth level of market share for partition k. In this way, the main formulation can be expressed as:

$$
\begin{aligned}
\textbf{(Main)} \qquad & \max_{x,z} \sum_{k} x_{i,j,k,v}\, \Pi_{i,j,v} && && (2.7)\\
& s.t. \sum_{i} x_{i,j,k,v} E_{i,j,v} = z_{j,k,v} \mathcal{Z}_j && \forall j,k,v\\
& \sum_{j,k,v} x_{i,j,k,v} = 1, && \forall i\\
& \sum_{j,v} z_{j,k,v} = 1, && \forall k\\
& \sum_{j,k,v} z_{j,k,v} P_v = 1,\\
& x_{i,j,k,v} \text{ binary}, \quad z_{j,k,v} \text{ binary}
\end{aligned}
$$

Compared to formulation (2.6), which selects one feasible column and a subset of rows (sellers) in the table, the main formulation draws entries from the tabulated tensor. Again, the first constraint ensures that the exponential of individual attrac-

tiveness sums to $\mathcal{Z}$. The second constraint guarantees that every seller is assigned to one and only one partition. The third constraint ensures that for each partition k, only one $\mathcal{Z}_j$ and one market share P_v are selected. Finally, we add the fourth constraint to ensure that the market shares assigned to each partition sum to one. Leveraging formulation (2.7), we showcase the optimal display policy and the corresponding optimal price in Figure 2-3. Specifically, Panel (a) of Figure 2-3 displays the equilibrium price and partitioning of each product under different demand levels. Panel (a) indicates that sellers set higher equilibrium prices when the platform faces higher demand. Interestingly, although the platform is provided with the flexibility to allocate traffic to all partitions, the optimal strategy turns out to be a subset display that only presents the medium to high-quality products to the customers. Meanwhile, as the demand increases, a larger set of products are displayed to the customers. In Panel (b), we require the demand allocated to each partition to be equal, and as a result, the subset partition is no longer feasible. Panel (b) demonstrates that there can be a significant imbalance in the number of products included and the average product quality among partitions depending on the specific demand level. We will address the concern of such potentially unfair allocation later in Section 2.5.

Finally, the main formulation can be solved by classic solvers such as Gurobi. Yet, the tabulation and optimization processes will become computationally challenging as the number of sellers N increases and step sizes of discretization of $\mathcal{Z}$ become finer. To reduce the computation time, we propose two relaxed formulations that provide the upper and lower bounds of formulation (2.7), respectively. The details of the simplification procedure and the goodness of approximation are discussed in Appendix A.2.

2.5 Fairness

While the partitioned display may increase the platform revenue compared to the display-everything strategy, the optimal partitions shown in Figure 2-3 can raise fairness concerns. First, the platform may display only a subset of products to achieve

higher profitability. While such a policy might be practical in offline retail settings (Heese and Martínez-de Albéniz 2018, Martínez-de Albéniz and Roels 2011), where the limited shelf space naturally prevents the store from showcasing all the products, this display policy potentially prevents a proportion of the sellers from joining the platform. It could spur concerns over an unfair business environment on online marketplaces, which the European Commission has regulated against (Bostoen 2018). Second, under the optimal display strategy, different partitions can vary substantially in the number of sellers and product quality. Such drastic variations in the assortment size and quality also raise whether such a display policy is fair to the customers. In this section, we address the unfairness in the display policy viewed by both sellers and customers by introducing several fairness metrics and incorporating them into formulation (2.7) as additional constraints.

2.5.1 Seller Fairness

We first study the fairness issue from the seller's viewpoint. Previous literature has proposed a number of commonly used notions of fairness (e.g., Mo and Walrand 2000, Bertsimas et al. 2011). For sellers, our goal is to ensure different partitions have equal access to the market and are similar in overall quality. As a result, our fairness definition resembles that in Cohen et al. (2019) and emphasizes the closeness in demand allocation and the partition competitiveness across different partitions. Specifically, each partition k is characterized by the incoming demand M_k and the total attractiveness in equilibrium $\mathcal{Z}^k = \sum_{\{i \in \mathcal{S}_k\}} \exp(a_i - \beta p_i^*)$. If two partitions k_1 and k_2 share the same values of $\mathcal{Z}^k$ and M_k, customers will face the same optimization problem and, therefore indifferent to the assignment. Motivated by this observation, we propose a notion of fairness measured by the relative difference in attractiveness $\mathcal{Z}$ and the difference in the incoming demand that each of the K partitions receives. To this end, we introduce the definition of (α, δ)-fairness.

Definition 1. *A display policy is (α, δ)-fair, where $0 \leq \alpha \leq 1$ and $0 \leq \delta \leq 1$, if the total attractiveness of each partition satisfies $\mathcal{Z}^k / \mathcal{Z}^{k'} \geq \alpha$ and the total demand*

assigned to each partition satisfies $M_k/M_{k'} \geq \delta$ *for all partitions* $k, k' \in \{1, ..., K\}$.

Intuitively, α quantifies the level of unbalance in the total attractiveness among partitions. According to the definition above, $\alpha = 1$ refers to the fairest scenario because each partition has the same total attractiveness, while a smaller α allows certain partitions to be significantly better or worse off than the rest and corresponds to a less fair scenario. Similarly, the demand allocation is fairest when $\delta = 1$, as the demand assigned to each partition is identical and equal to M/K in this case. When $\delta = 0$, the fairness constraint ceases to matter, and the platform can freely allocate the market shares among partitions. In fact, the notion of (α, δ)-fairness could be easily incorporated into our MIP Formulation (2.7) as additional constraints. Denoting the feasible region of Formulation (2.7) by $\mathcal{Q}$, we augment Formulation (2.7) to include the α and δ fairness constraints as

$$
\begin{aligned}
\max_{x,z} \quad & \sum_{k} x_{i,j,k,v}\, \Pi_{i,j,v} && && (2.8)\\
s.t. \quad & \sum_{j,v} z_{j,k,v} \mathcal{Z}_j \geq \alpha \sum_{j,v} z_{j,k',v} \mathcal{Z}_j && \forall k, k' = 1, 2, \ldots, K \\
& \sum_{j,v} z_{j,k,v} P_v \geq \delta \sum_{j,v} z_{j,k',v} P_v && \forall k, k' = 1, 2, \ldots, K \\
& x, z \in \mathcal{Q}
\end{aligned}
$$

The integration of fairness into the main formulation is straightforward. Recall that $\mathcal{Z}_j$ is the j-th level of the tabulated $\mathcal{Z}$ value and P_v is the v-th level of the tabulated market share. The first two constraints ensure that the ratios of the market share and attractiveness between any two partitions are greater than δ and α, respectively. Moreover, note that Formulation (2.8) does not guarantee the existence of a feasible solution. In other words, Formulation (2.8) might be infeasible if we impose somewhat restrictive fairness constraints, i.e., both δ and α are required to be close to one.

The (α, δ)-fairness guarantees a certain degree of equity in demand and attractiveness across partitions, but it is unclear how an (α, δ)-fair policy would affect every single seller's revenue. To quantify the effect of various fairness policies from the per-

spective of an individual seller's revenue, we propose the following definition of the "envy level".

Definition 2. *Denote by Π_i^0 the revenue that seller i receives from the current partition. When seller i is moved to the k-th partition, $k = 1, 2, \ldots, K$, while the rest of the sellers remain unchanged, we denote the revenue that seller i will receive by Π_i^k. Then, we define the current "envy level" for seller i as:*

$$EN_i = \frac{\max_k\{\Pi_i^k\} - \Pi_i^0}{\Pi_i^0}.$$

As a measure of the potential percentage of revenue gain, the envy level captures the incentive for one seller to move to another partition. A display policy is considered less fair if the envy level is large for some sellers under such policy, as these sellers would have strong incentives to move to a more profitable partition. On the other hand, if $EN_i = 0$ for all $i = 1, 2, \ldots, N$, we consider the current partition fair because no seller has incentives to switch partition groups unilaterally. Notably, when a display policy is (α, δ)-fair, i.e., $\delta = \alpha = 1$, we also have $EN_i = 0$ for all $i = 1, 2, \ldots, N$, indicating that sellers would be indifferent to which partition they are assigned, as their equilibrium prices and revenues will be exactly the same across all partitions.

Exploiting Formulation (2.8), we are able to derive the equilibrium price and the resulting seller revenue under different (α, δ) combinations. In Panel (a) of Figure 2-4, we demonstrate how the maximum of the envy level across sellers varies under different (α, δ) combinations. As expected, the envy level is the highest when the display policy is the least fair as measured by α- and δ-fairness, i.e., $(\alpha, \delta) = (0, 0.2)$. Moreover, Panel (b) of Figure 2-4 showcases how the total revenue as defined in the objective of Formulation 2.8 is affected by the display fairness. We observe that the total revenue is monotonically decreasing as either α or δ increases. The revenue under the fairest case in which $(\alpha, \delta) = (1, 1)$ is 86.1% of that when the platform is not concerned about fairness at all, i.e., $(\alpha, \delta) = (0, 0)$. Additionally, we scrutinize the relationship between seller quality and each seller's envy level. Specifically, we

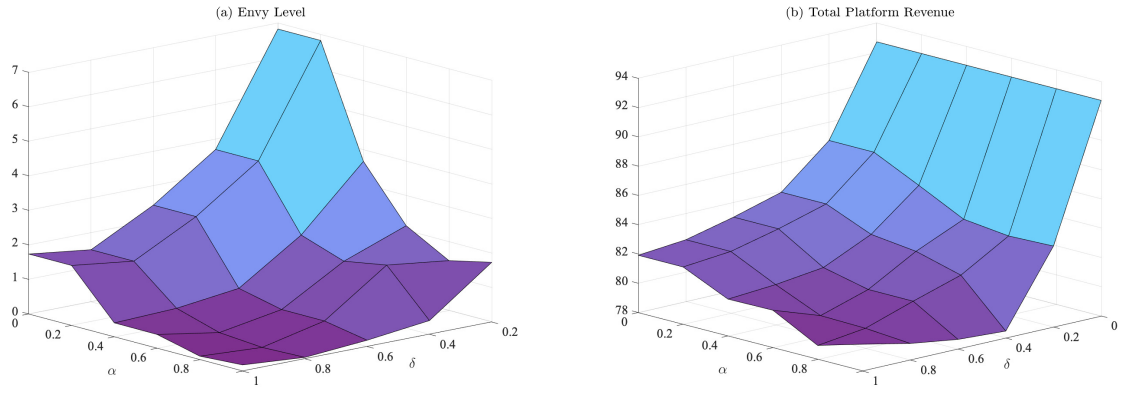

Figure 2-4: Envy Level and Total Platform Revenue under Different α and δ. The full assortment has 40 sellers with quality a_i following normal distribution $\mathcal{N}(4, 1)$. Demand is set to be 60, and the partition number is set to be 3. Parameters α and δ are set from 0 to 1 with intervals of 0.2.

sort 40 sellers according to their product quality, classify them into three groups with sizes of 13, 13 and 14 and label them as 'Low,' 'Medium,' and 'High,' respectively. We plot the envy distribution of these three groups in Figure 2-5. Notably, sellers in the low-quality group exhibit a significantly higher envy level than their counterparts in the higher quality groups, as shown in Figure 2-5, indicating that a seller with a lower quality product is more likely to receive an unfair assignment. In fact, the maximal envy level of the 'High' group is 0.58, while for the 'Low' group, it is as high as 6.85. Without resorting to Formulation (2.8), we can use the following proposition to bound the envy level of each individual seller.

Proposition 2.3. *For any display policies with fairness level* $\frac{2}{3} < \alpha \leq 1$ *and* $0 < \delta \leq 1$*, the envy level for each seller can be bounded by*

$$EN_i \leq \frac{2\alpha}{3\alpha - 2}\left(\frac{1}{\alpha} + \frac{1}{\delta} - 2\right), \quad \forall i = 1, 2, \ldots, N. \tag{2.9}$$

This proposition aims to illustrate the connection between the individual envy level and fairness parameters (α, δ). The upper bound of the envy level is decreasing in both δ and α. Moreover, the rate of change is significantly higher when δ and α are small. This indicates that subject to the same decrease in the fairness parameters, the envy level is likely to increase in less fair settings. Although the left-hand side

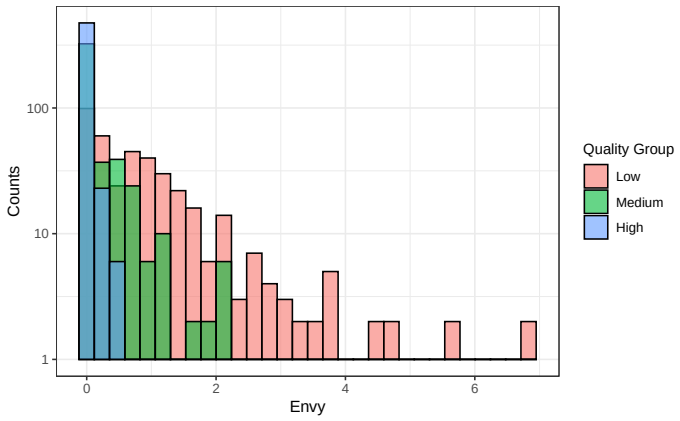

Figure 2-5: Envy Distribution for Different Quality Groups. The parameter setup is the same as that in Figure 2-4.

EN_i is on individual sellers, the right-hand side upper bound is free of any individual inputs. In other words, the bound is tightest when benchmarked with the seller that has the maximum envy level.

2.5.2 Customer Fairness

Similar to how the profit-maximizing partition strategy can cause unfair distribution of exposure among sellers, the variations in assortment size and quality across partitions can also lead to the perception of unfairness among customers. Meanwhile, it is unclear if a display policy that is fair to the sellers is also fair for the customers. To address these questions, we first explore the connections between the seller and customer fair. We then formally define and formulate customer fairness as additional constraints to formulation 2.7.

We define the welfare a customer in partition k receives as u^k. As all customers are homogeneous in our model, u^k's are independent and identically distributed (i.i.d.) across customers in the same partition. As a result, we drop the customer subscript from the notation. Specifically, u^k equals $a_i - \beta p_i^*$ if the customer books seller i's product, and zero if the customer selects the outside option. To investigate to what extent seller fairness is equivalent to customer fairness, we study through numerical analysis the difference in the expected individual customer welfare among customers

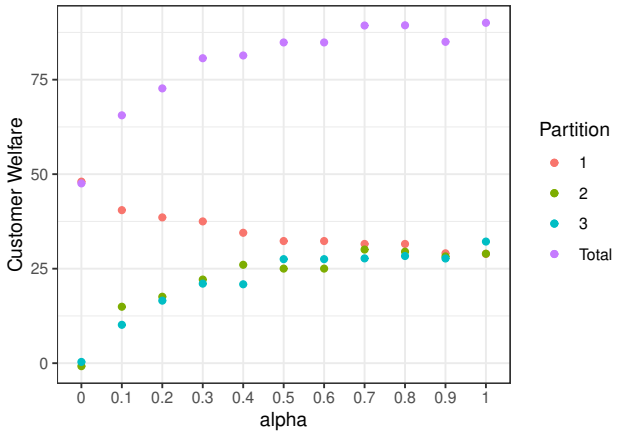

Figure 2-6: Customer Welfare under Different Level of alpha in the Equal Demand Case.

in three partitions under $(\alpha, 1)$-fair display policies. In this case, we consider customer fairness as the relative difference in customer welfare across partitions. We plot the equilibrium outcomes in Figure 2-6. When we impose more strict fairness requirements on the sellers, i.e., when α approaches one, the differences in customer welfare across partitions shrink and eventually evaporate. This numerical study suggests that equity in customer welfare across partitions can be achieved with a display policy that is $(1, 1)$-fair to the sellers.

Figure 2-6 suggests that a certain degree of customer fairness can be simultaneously guaranteed when the platform imposes restrictions on seller fairness. Nevertheless, the converse of the statement that customer fairness leads to seller fairness does not always hold. To see this, consider a subset display policy that allocates all the demand into one partition. This policy is fair to all customers since they are assigned to the same partition. However, such display is unfair to sellers who are not assigned any demand. As a result, we need to develop a separate notion of fairness for the customers. Similar to Definition 1, we define customer fairness as follows:

Definition 3. *A display policy is γ-fair, where $0 \leq \gamma \leq 1$, if the expected individual customer welfare of each partition satisfies $\mathbb{E}[u^k]/\mathbb{E}[u^{k'}] \geq \gamma$ for all partitions $k, k' \in \{1, ..., K\}$.*

The parameter γ reflects the closeness of expected individual customer welfare

across partitions. Again, while we define seller fairness on the partition level, customer fairness can be interpreted as either a partition level or an individual level of fairness metric since all the customers are assumed to be homogeneous in our model.

We next incorporate the notion of customer fairness into our MIP formulation. In this case, we are not able to directly tabulate the expected customer welfare $\mathbb{E}[u^k]$ as described in Algorithm 1. Previously, the tabulation of FOC as shown in Equation (2.4) is possible because we manage to decompose the system of equations to a state in which the solution p_i^* depends only on seller i's product quality a_i and the intermediate variable $\mathcal{Z}$ so that we can precompute and store the solution of each individual equation. The computation for $\mathbb{E}[u^k]$, however, requires enumerating all the demand realizations under all potential partition outcomes, which is a combinatorial problem by nature.

Given that it is infeasible to compute $\mathbb{E}[u^k]$, the expected welfare that customers in partition k receive, we propose to circumvent this challenge by tabulating how much customer welfare each individual seller generates in expectation. Intuitively, the equivalence of these two terms can be justified by drawing an analogy from the network flow problem: the total customer welfare that customers in partition k receive must be equal to the total amount generated by the sellers in partition k. We denote the customer welfare generated by seller i in partition k as v_i^k. There are two realizations for v_i^k: $v_i^k = a_i - \beta p_i^*$ if the product is purchased, and $v_i^k = 0$ if no customer ends up purchasing the product. The purchasing probability of product i in partition k is described in Equation (2.3). In this way, we can compute $\mathbb{E}[u^k]$ through the following equation

$$\mathbb{E}[u^k] = \frac{1}{M_k}\sum_{i\in\mathcal{S}_k}\mathbb{E}[v_i^k] = \frac{1}{M_k}\sum_{i\in\mathcal{S}_k}(a_i-\beta p_i^*)\left(1-\left(1-\frac{\exp\left(a_i-\beta p_i^*\right)}{1+\sum_{j\in\mathcal{S}_k}\exp(a_j-\beta p_j^*)}\right)^{M_k}\right). \tag{2.10}$$

Now we can tabulate the expected customer welfare seller i generates under jth level of $\mathcal{Z}$ and market share level v as

$$U_{i,j,v} = \frac{1}{MP_v}(a_i-\beta p_{i,j,v}^*)\left(1-\left(1-\frac{\exp\left(a_i-\beta p_{i,j,v}^*\right)}{1+\mathcal{Z}_j}\right)^{MP_v}\right). \tag{2.11}$$

As the tabulated $U_{i,j,v}$'s are seller dependent, we need to sum over i, j, v to derive the partition-level customer welfare. When the platform requires the display policy to be γ-fair, we can formulate the problem as

$$\begin{aligned} \max_{x,z} \quad & \sum_{k} x_{i,j,k,v}\, \Pi_{i,j,v} && (2.12) \\ s.t. \quad & \sum_{i,j,v} x_{i,j,k_1,v} U_{i,j,v} \geq \gamma \sum_{i,j,v} x_{i,j,k_2,v} U_{i,j,v} && \forall k_1, k_2 = 1, 2, \ldots, K \\ & x, z \in \mathcal{Q}\,. \end{aligned}$$

2.6 Application using Airbnb Datasets

This section presents a case study using transaction data from Airbnb to demonstrate how our framework can be applied in practice. Our goal is to show the empirical strategy when fitting our model to the data and ultimately evaluate the impact of different display policies under various market conditions for the Airbnb platform. Our model, albeit stylized, closely reflects the gist of the decision-making processes of the hosts and the Airbnb platform for the following reasons. First, each host on the Airbnb platform supplies a listing with unit availability every day. Second, many hosts have disabled instant booking, indicating that they plan to evaluate the profiles of all the applicants before accepting one, thus translating the sequential arrival pattern of the customers to a simultaneous scenario. Finally, each host makes independent pricing decisions while observing the prices set by other listing owners.

In what follows, we first introduce the setting of Airbnb and provide a summary of our data. We then describe how to fit our model to the transaction data to estimate a_i, the quality of each Airbnb listing i. Finally, we tabulate the critical inputs of our MIP, through which we examine the optimal display policy under different demand scenarios.

2.6.1 Data

Our Airbnb dataset covers the transaction history of listings in Manhattan, New York, in 2018. Airbnb offers different home types that include Private Room, Shared Room, and Entire Home/Apartment. Different home types tend to target different customer segments, so competition usually occurs only within each home type. Thus, we limit the scope of our analysis to include only listings labeled as Entire Home/Apartment. The data include the daily transaction history of 2,561 such listings. On each day, the data documents the status of each listing as either blocked (i.e., made unavailable by the owner), available, or reserved. This way, we can recover the assortment of listings displayed to the customers that we use to form the consideration sets for the customers. The data also provide daily booking prices for each listing and other listing characteristics such as overall ratings and number of reviews. We present the summary of the listing characteristics in Table 2.2.

Table 2.2: Summary of Listings Characteristics

Property characteristics	Mean	St. Dev.	Min	Max
Occupancy Rate	0.65	0.36	0	1
Price (in USD)	247.54	287.58	10	2,500
Number of Reviews	70.02	63.49	0	400
Overall Rating	4.59	0.61	0	5
Number of Bedrooms	1.24	0.89	0	6
Number of Bathrooms	1.13	0.41	0	5
Response Rate†	91.19	19.12	0	100
Superhost	0.23	0.42	0	1

† Note: Response rate is defined as the percentage of the time that a host responds to guests within 24 hours. A host becomes as a Superhost if the host satisfies a series of criteria set by Airbnb, such as a high overall rating and low cancellation rate.

2.6.2 Estimation

To derive the optimal display policy for Airbnb, we first need to estimate each listing's quality. According to our theoretical model in Section 2.3, the overall mean utility of listing i on day t can be expressed by the listing quality and price and as $a_i - \beta p_{it}$. Nevertheless, instead of fitting an individual a_i for each listing, we parametrize listing quality as a linear combination of listing-specific covariates to reduce the number of

parameters to be estimated and avoid overfitting the model. Specifically, we write the utility that customer m gains from booking listing i on day t as

$$\mu_{itm} = a_i - \beta p_{it} + \epsilon_{im} = \boldsymbol{X_i}\boldsymbol{\gamma} + FE_i - \beta p_{it} + \epsilon_{im}\,,$$

where ϵ_{im} follows i.i.d. $Gumbel(0,1)$. Quality is expressed as $a_i = \boldsymbol{X_i}\boldsymbol{\gamma} + FE_i$, where $\boldsymbol{X_i}$ contains the listing characteristics summarized in Table 3.1 (apart from occupancy rate and price). Additionally, following Li et al. (2019), we divide Manhattan into 10 regions, and assume that substitution occurs within each region. We use FE_i to represent the regional fixed effect that listing i shares with competing listings within the same region. For ease of notation, we write $\boldsymbol{\theta\Xi_{it}} = \boldsymbol{X_i}\boldsymbol{\gamma} + FE_i - \beta p_{it}$. In this way, the theoretical probability that listing i is booked on day t is

$$1 - (1 - q_{it})^M = 1 - \Big(1 - \frac{\exp(\boldsymbol{\theta\Xi_{it}})}{1 + \sum_{j \in \mathcal{S}_{it}} \exp(\boldsymbol{\theta\Xi_{it}})}\Big)^M\,, \tag{2.13}$$

where $\mathcal{S}_{i,t}$ is the partition that listing i belongs to on day t. Since Airbnb by default displays all the available listings over a certain region to all customers, $\mathcal{S}_{i,t}$ in this case represents all the listings within the same neighborhood as listing i. We denote by the binary variable $Y_{i,t}$ the observed outcome from the transaction history indicating whether listing i is booked on day t. We adopt the nonlinear least squares framework to recover a consistent estimate of $\boldsymbol{\theta}$ that leads to the best fit of the demand realization:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{t=1,\ldots,T} \sum_{i=1,\ldots,N} \left(1 - \Big(1 - \frac{\exp\left(\boldsymbol{\theta\Xi_{it}}\right)}{1 + \sum_{j \in \mathcal{S}_{i,t}} \exp\left(\boldsymbol{\theta\Xi_{jt}}\right)}\Big)^M - Y_{it}\right)^2. \tag{2.14}$$

We would like to make several notes about this estimation equation. First, the listing-specific parameters are identifiable because we observe variations in these listing characteristics in the data. Moreover, the identifiability of the regional fixed effects is achieved through the variations in the overall occupancy rate across different regions. One estimation challenge is that Equation (2.14) requires the actual demand level M as a fixed input, which in reality reflects listing owners' common belief about future

Table 2.3: Estimation Results

	Demand/Supply Ratio		
	$(M/N = 1)$	$(M/N = 1.5)$	$(M/N = 2)$
Price ($\times$ 100)	-0.193	-0.217	-0.291
	(-0.202 -0.185)	(-0.235 -0.205)	(-0.391 -0.25)
Overall Rating	0.355	0.382	0.499
	(0.332 0.378)	(0.355 0.413)	(0.410 0.691)
Number of Reviews	0.005	0.006	0.011
	(0.005 0.005)	(0.006 0.006)	(0.009 0.012)
Superhost	0.100	0.124	0.184
	(0.078 0.121)	(0.097 0.153)	(0.122 0.231)
Response Rate	0.007	0.007	0.007
	(0.006 0.007)	(0.006 0.008)	(0.006 0.008)
Number of Bedrooms	0.174	0.197	0.244
	(0.163 0.187)	(0.181 0.220)	(0.214 0.349)
Number of Bathrooms	0.205	0.231	0.311
	(0.179 0.232)	(0.191 0.304)	(0.232 0.425)
Observations	46,769	46,769	46,769
Regional Fixed Effects	Yes	Yes	Yes

demand. However, we do not directly observe M in the data, so we conduct estimation under different demand scenarios. Specifically, we assume M to be proportional to the number of listings, with the demand-to-supply ratio set to be 1.0, 1.5 or 2.0. The estimation results are presented in Table 2.3.

As the demand level increases, Table 2.3 suggests that while the magnitude of these estimates increases, the relative magnitude across estimates largely persists within each demand scenario. The 95% confidence intervals of our estimation results obtained through bootstrap simulations are presented in brackets. Using the estimation results, we calculate the listing quality as $a_i = \boldsymbol{X_i}\boldsymbol{\gamma} + FE_i$.

2.6.3 Counterfactual Analysis

We obtain three sets of listing quality using our estimation results, each corresponding to a demand level. For each demand scenario, we incorporate the listing quality into Algorithm 1 to tabulate $p^*_{i,j}$, $E_{i,j}$ and $\Pi_{i,j}$, which in turn are used as inputs for our MIP formulation, i.e., Equation (2.7). We can provide recommendations on the optimal partition number for each neighborhood under each demand scenario. We visualize the solutions from our MIP in Figure 2-7.

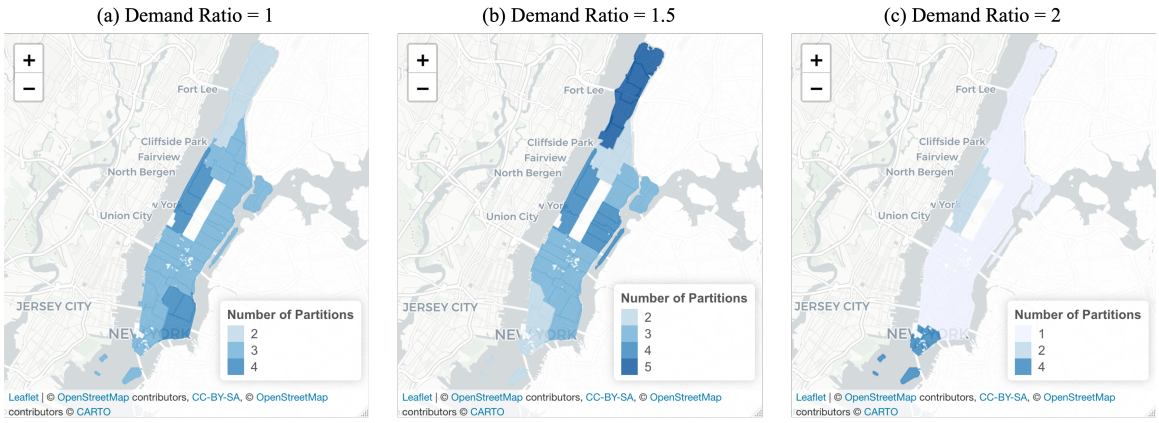

Figure 2-7: Optimal Number of Partitions in each Manhattan Neighborhood.

When listing owners expect the daily demand to be low or moderate, Panels (a) and (b) suggest that Airbnb should assign listings in each neighborhood to partitions. Nevertheless, Panel (c) of Figure 2-7 implies that Airbnb should use the current display-everything strategy in most neighborhoods when faced with sufficient demand, which is consistent with Theorem 2.1. Importantly, in addition to the optimal partition number, our MIP formulation also explicitly indicates which partition each listing is assigned to, making our results readily implementable by the platform.

Additionally, we also investigate the cost of fairness in the context of Airbnb. To this end, we compare the difference in the optimal partition number and revenue when the display policy is $(0, 0)$-fair and $(0, 1)$-fair. We present the results of our counterfactual analysis in Table 2.4. Under each demand scenario, we report the optimal partition numbers under the two abovementioned policies in parentheses. The resulting revenue gap generally falls within 20%. Notably, when demand is sufficiently high, the unconstrained optimal display policy is also $(0, 1)$-fair for most of the neighborhoods, as the optimal strategy is to display all listings to all the customers, which automatically satisfies the definition of $(0, 1)$-fairness. Using a similar approach, we can assess the profit gap between the unconstrained optimal display policy and the optimal policy under arbitrary α, δ, and γ combination.

Table 2.4: Optimal Partition Number (PN) when $(\alpha, \delta) = (0, 0)$ and $(0, 1)$ and the Revenue Gap.

District	$DR = 1.0$		$DR = 1.5$		$DR = 2.0$	
	PN	Gap	PN	Gap	PN	Gap
Central Harlem	(3, 3)	18.12%	(2, 4)	17.23%	(1, 1)	0.00%
Chelsea and Clinton	(3, 5)	19.84%	(3, 2)	21.84%	(1, 1)	0.00%
East Harlem	(3, 5)	15.53%	(3, 2)	12.84%	(1, 1)	0.00%
Gramercy Park and Murray Hill	(3, 5)	18.72%	(3, 2)	19.65%	(1, 1)	0.00%
Greenwich Village and Soho	(4, 5)	19.65%	(3, 4)	17.60%	(1, 1)	0.00%
Inwood and Washington Heights	(3, 5)	18.15%	(2, 3)	19.23%	(1, 1)	0.00%
Lower East Side	(3, 2)	9.26%	(2, 1)	14.14%	(4, 1)	55.02%
Lower Manhattan	(3, 4)	19.35%	(4, 3)	18.28%	(1, 1)	0.00%
Upper East Side	(4, 4)	17.76%	(4, 4)	16.22%	(2, 1)	5.79%
Upper West Side	(2, 3)	15.42%	(5, 4)	14.55%	(1, 1)	0.00%

In fact, Airbnb has been segmenting the listings into different partitions since launching 'Airbnb Plus,' which, according to Airbnb, is a selection of only the highest quality homes with hosts known for great reviews and attention to detail. Our partition display policy can be implemented in a similar fashion. Customers searching for lodging options in a neighborhood will be presented with a selection of listings, with the probability of seeing each partition governed by the market share allocated to each partition.

2.7 Extension to Non-Unit Inventory

Unlike Airbnb, sellers on platforms such as Amazon or Taobao may possess more than one unit of inventory for each product. We thus formulate the objective functions and derive the corresponding FOCs under multi-unit and infinite-unit scenarios. We theoretically derive the optimal display policy when the platform is sufficiently large. Notably, in the non-unit inventory case, we can still derive the optimal display pol-

icy under arbitrary market conditions and fairness constraints by incorporating the outputs from the tabulation procedure described in Table 1 into the MIP formulation.

2.7.1 Finite Inventory

First, we assume that seller i owns one SKU and has W_i units of inventory in stock. We assume the market size to be larger than the inventory level, i.e., $W_i < M$. The seller's objective function when facing demand M can be written as

$$\max_{p_i} \Pi_i(M, p_{-i}) = p_i \left(W_i - \sum_{j=0}^{W_i-1} (W_i - j) \binom{M}{j} q_i^j (1 - q_i)^{M-j} \right) \tag{2.15}$$

where q_i is defined as in Equation (2.2). Note that, instead of expressing the total revenue as the sum of revenue from selling $1, 2, ...W_i$ products, Equation (2.15) captures the difference in the total revenue collected from selling all the products and sum of the revenue losses when there are $1, 2, ...W_i$ unsold products. Then, the FOC of Equation (2.15) can be expressed as

$$\sum_{j=0}^{W_i-1} (W_i - j) \binom{M}{j} q_i^j (1 - q_i)^{M-j} (1 + M\beta p_i q_i - j\beta p_i) = W_i. \tag{2.16}$$

Given an arbitrary M, obtaining the closed-form solution for Equation (2.16) is analytically challenging due to the combinatorial and nonlinear nature of the equation. Nonetheless, when demand M approaches infinity, the probability that all the W_i products are purchased approaches 1. Thus, Equation (2.16) reduces to $(1+M\beta p_i q_i - j\beta p_i) = W_i$, which allows us to obtain the following result:

Theorem 2.2. *Suppose that there are N products $\mathcal{S} = \{1, 2, \ldots, N\}$, each with finite inventory W_i. For any $\gamma > 1$, there exists a threshold $M(\gamma)$, such that when $M > M(\gamma)$, for any display policy $\{\mathcal{S}_k, M_k\}_{k=1}^K$ that satisfies $M_k < M/\gamma$, we have*

$$\sum_{i=1}^{N} \Pi_i^*(\mathcal{S}, M) > \sum_{k=1}^{K} \sum_{i \in \mathcal{S}_k} \Pi_i^*(\mathcal{S}_k, M_k).$$

Thus, when each product has finite units, it is still optimal for the platform to display the entire assortment to all customers when demand M is sufficiently large.

2.7.2 Infinite Inventory

Finally, we consider an extreme case in which each vendor on the platform holds infinite inventory for the listed product. In this case, the seller's objective function can be written as:

$$\max_{p_i} \Pi_i^s(p_{-i}) = p_i q_i = \frac{p_i \exp(a_i - \beta p_i)}{1 + \sum_1^N \exp(a_j - \beta p_j)}. \tag{2.17}$$

When all the demand can be satisfied, the purchasing probability in Equation (2.17) reduces to the standard MNL model (which is also equivalent to Equation (2.7) by setting $M = 1$). Then, the FOC for Equation (2.17) becomes

$$\frac{1}{\beta p_i} = 1 - q_i \tag{2.18}$$

In this case, the equilibrium price and revenue no longer depend on total demand. Thus, the platform's problem reduces to deciding only the number of listings in each partition. Intuitively, when the platform hosts many sellers, with each possessing a high level of inventory, it can be optimal to display only a subset of products. Proposition 2.4 formally characterizes the optimal display policy when the number of sellers is large:

Proposition 2.4. *Denote by $\Pi(\mathcal{S})$ the total revenue from displaying assortment $\mathcal{S}$ to customers. Suppose that product quality satisfies $a_{ub} = a_1 \geq a_2 \geq \cdots \geq a_N = a_{lb}$, where $a_{lb} > 0, a_{ub} > 1/\beta$ and $a_{ub} - a_{lb} < \ln(\beta a_{ub} - 1)$. Then, there exists a threshold N_0 such that when the cardinality of $\mathcal{S}$ satisfies $|\mathcal{S}| > N_0$, we have $\Pi(\mathcal{S}) < \Pi(\{a_1\})$.*

Theorem 2.4 indicates that when the market becomes sufficiently competitive, and the products are similar in quality, the platform should only display the product with the highest quality. Notably, our numerical results of the unit-inventory case presented in Panel (b) of Figure 2-2 point to a similar conclusion: when the number

of products offered is considerably more significant than the market size; it is optimal to display only a small subset of the entire assortment.

Chapter 3

Estimating and Exploiting the Impact of Photo Layout: A Structural Approach

3.1 Introduction

Airbnb is a peer-to-peer lodging marketplace provider that offers close to 5 million listings across 81,000 cities and has hosted over 300 million guests since 2008. The platform aims to empower each of its hosts to provide high-quality rentals and maximize their revenues. One of Airbnb's challenges since as early as 2011 is to make sure that images posted by property owners are captivating and properly presented. However, unlike the hotel industry, where hotels take photos of their rooms, Airbnb has little control over user-generated images. The platform launched a photography program in 2011 to help match hosts with local professional photographers who help take photos of the hosts' properties to improve photo quality. Although Airbnb claims that images taken by professional photographers can lead to a 40% higher total earnings, 24% more bookings, and 26% increased nightly prices[1], many hosts still post low-quality photos (Zhang et al. 2019), let alone perform a full-fledged optimization

[1]https://www.airbnb.com/professional_photography.

of photo layout that accounts for image content, image quality, and the display order.

There is a good deal of literature that demonstrates the impact of images on consumer behaviors within advertisement settings (Meyers-Levy and Peracchio 1992, Miller and Kahn 2005). Similarly, Airbnb hosts use images to reveal critical information about their apartments. Since customers usually have limited time and attention when booking their property, they rely heavily on visual information to quickly compare alternatives. Based on the quality and content of the images, customers form expectations about each property and accordingly make decisions about which lodging to choose. Thus, given a collection of images with varying quality and content, deciding which photo to prioritize on the listing's web page is crucial for attracting customers. Building upon previous papers that study the impact of various factors on property demand (Li et al. 2016, Zhang et al. 2017), we examine how customers' renting decisions are affected by photo layouts, which are characterized by the room type featured in the photo, the photo quality and the order of display on the listings' webpages.

For this study, we use three datasets to conduct our empirical analysis. The first dataset contains detailed property characteristics for the 10,280 listings on Airbnb in New York City in 2018. The second dataset documents the daily transaction history for each listing throughout the year. The third dataset contains more than 220,000 photos posted by the apartment owners on the Airbnb platform for the same set of listings. Because information on photo quality and photo content is not readily available in the existing dataset, we employ techniques in Computer Vision to augment our information on image quality and image room types. Specifically, to extract image quality and image room types from each photo, we apply a widely-used convolutional neural network (CNN) model, Resnet50 (He et al. 2016), and we build two separate supervised learning models to evaluate image quality and classify photo type. A score from 1 to 7 is assigned as a measurement for photo quality, where a higher score indicates that a photo is more visually attractive. Meanwhile, photos are classified into one of the five room categories, namely, *BEDROOM*, *LIVING ROOM*, *OUTSIDE*, *TOILET* and *KITCHEN*. The accuracy of the image quality scoring model and the

room type classification model are 85.7% and 84.0%, respectively.

To examine and maximize the impact of a listing's photo layout on its demand using the three datasets, we address three major challenges: (i) the definition and the identification of layout need to be carefully specified and carried out; (ii) as each listing on each day can only be booked once, the demand data will be severely censored to 1 and customers would face changing choice sets, which make the estimation results from traditional estimation approaches such as the multinomial logit (MNL) model biased in the Airbnb setting; and (iii) posting photos of the same room type may strengthen or weaken the impact from each individual photo. Thus the formulation and the solution of the layout optimization problem are non-trivial.

We quantify the overall impact of photo layout according to how these photos are displayed on the Airbnb website. As shown in Figure 3-1, customers can view thumbnails of multiple listings within the search page. If they are interested, consumers can proceed to any listing page. By default, the listing's page displays the first five images of a listing, in which the cover image is emphasized and takes up a more significant amount of space than the other four photos. To reflect how Airbnb displays listing images, we let the impact of photo layout depend on each photo's quality and room type. More importantly, whether an image is posted in the cover spot versus the other four spots. We also introduce duplicate factors to capture the increased or dampened marginal benefits from posting multiple images that contain the same room type. With the photo-level data, we conduct reduced-form analysis to study how photo layout affects the monthly occupancy rate of each property. We observe that higher quality images lead to a higher occupancy rate. In addition, we find that the cover image generally has a more significant impact on a listing's occupancy rate than non-cover images. The bedroom photo has a more substantial impact than other room types in the cover spot.

Next, we estimate the impact of image layout on customers' renting decisions under a random utility framework to pave the way for optimization and counterfactual analysis over image layout. Because each lodging provides only one unit of supply each day, the demand for each listing on each day is censored to 1. The censored

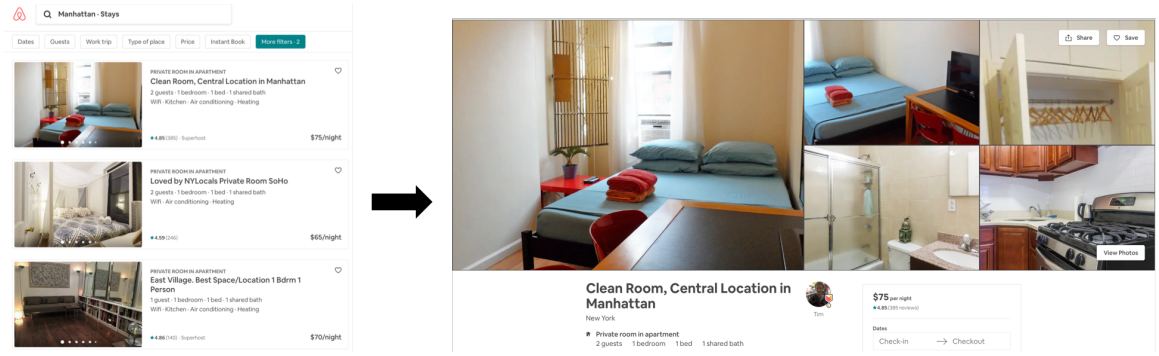

Figure 3-1: Screenshot of Airbnb Webpages. Left: Cover Thumbnails on the Search Page; Right: Listing View Page

demand data will bias the estimation in two ways: (i) The demand of the listings is truncated. Consequently, we cannot distinguish the relative attractiveness among popular listings that get reserved every day, and (ii) Once an apartment is booked for a specific check-in date, customers who visit the website later would face a smaller consideration set for the same check-in date. In this paper, we propose a novel pairwise comparison model to estimate customers' responses to the layout of photos consistently. We further show that the estimation process can be simplified under specific error term structures. In addition to sales data, which classic choice models usually use, the pairwise comparison model also utilizes information on the sequence by which each property is booked to recover the impact of property characteristics. The estimation results suggest that the cover image has a much higher impact on customers' perceptions of lodging quality than photos in the non-cover spots. We also observe that using a high-quality bedroom photo for the cover image results in the maximal increase of a listing's attractiveness. Interestingly, we do not find that living room cover photos, which more than 60% of the listings are currently using, make listings more attractive than the other room types. In addition, we observe decreasing (increasing) marginal returns in image quality from posting bedroom (living room) images in both cover and non-cover spots, suggesting listing owners should consider such duplicate effects when setting up their photo layouts.

The results from this paper should provide direction for an Airbnb listing owner when deciding which five photos to select from an existing collection of images to

maximize his listing's attractiveness. To this end, we formulate a non-linear integer programming optimization problem and develop an algorithm to derive the optimal photo layout. Under the optimal photo layout, 71% of listings use a bedroom image as their cover photo. The rest of the room types are featured as cover images only when they have much higher quality than the bedroom photos. We demonstrate through our counterfactual that when a listing unilaterally switches to the optimal photo layout, on average, the listing owner enjoys an 11.0% increase in booking probability. Additionally, suppose the listing owner increases its rental price to neutralize the benefits generated by the optimal photo layouts (i.e., the listing's rental probability stays the same). In that case, the yearly revenue will increase by an average of $1,248. Finally, when all listings on Airbnb simultaneously switch to the optimal photo layout, depending on the specific neighborhood and market size, each listing on Airbnb will be booked for two to five more days in a year, which boosts the revenue by $500 to $1100, respectively.

Our paper contributes to the OM community in the following two ways. First, our paper provides an integrated framework that extracts relevant photo features and quantifies and estimates the impact of the photo layout. This empirical strategy echoes and builds upon previous literature ((Martinez-de Albeniz and Valdivia 2019), (Zhang et al. 2019)) and can be used as an intermediate step for full-scale assortment optimization on the product display. Second, the PCM approach we propose can obtain less biased estimates in settings where the consideration set changes dynamically over time. Besides the Airbnb platform, such settings include the flash sales industry ((Boada-Collado and Martínez-de Albéniz 2020)), car rental industry ((Jagabathula et al. 2019)) and matching market ((Kanoria and Saban 2020)).

3.1.1 Literature Review

Our research is related to four streams of literature: the impact of photo layout on product sales, computer vision, consumer choice models and choice model-based assortment optimization, and the sharing economy, with a particular focus on peer-to-peer lodging rental platforms.

Using images as an effective means to convey information has been widely studied across various fields such as psychology and marketing. Valdez and Mehrabian (1994) and Mikels et al. (2005) show that images can affect people on an emotional level. About the advertisement industry, Snyder and DeBono (1985) suggests that manipulating products' images helps capture different types of customers. Moreover, previous research (() (Gorn et al. 1997), (Miller and Kahn 2005)) documents that image details such as camera angles and colors significantly impact images on consumer behaviors. Our paper differs from existing literature by jointly examining the impact of image quality, orders of image display, and image content on consumer choice in the peer-to-peer apartment-sharing market.

Extracting and evaluating information from images are classic computer vision tasks that have been well-studied by the machine learning community. Datta et al. (2006) is among the first to extract detailed image features such as saturation and hue and apply Support Vector Machine to predict binary high/low image quality outcomes. In recent years, the Convolutional Neural Network (CNN) has been introduced and shown to significantly improved the out-of-sample prediction accuracy over traditional techniques ((Krizhevsky et al. 2012), (Simonyan and Zisserman 2014)). In our paper, to evaluate the quality and content of images posted by Airbnb listing owners, we adopt Resnet50 - a type of pre-trained, structured CNN model capable of regression and classification computer vision tasks. As a 50-layer residual network, Resnet50 has been proven to outperform average human judgment on the ImageNet dataset ((He et al. 2016)).

Our paper is closely related to literature that studies choice-based demand models and their variants. The multinomial logit (MNL) model is one of the most frequently used and studied discrete choice models ((McFadden 1978), (Ben-Akiva and Lerman 1985)). To relax the assumption that the error term in the MNL model follows a general extreme value (GEV) distribution, Mahajan and Van Ryzin (2001) and Farias et al. (2013) introduce non-parametric rank-based choice models that focus on the preference list of the choices without specifying any utility forms. In addition, Manski (1975) and Fox (2007) develop semi-parametric choice models that assume

flexible error structures. In our paper, we must deal with two unique challenges in the Airbnb setting that render classic estimation approaches non-viable, namely shrinking consideration sets and heavily censored demand ((Boada-Collado and Martínez-de Albéniz 2020)). In response to these challenges, we develop a pairwise comparison estimation model (PCM) that utilizes information about the sequence with which each property is booked to consistently estimate the impact of photo layouts on customers booking decisions. In addition, PCM does not require a specific error term structure and is more robust to error term misspecification.

We also contribute to literature that examines choice model-based assortment optimization problems (Mahajan and Van Ryzin 2001, Talluri and Van Ryzin 2004, Kök and Fisher 2007, Aouad et al. 2015). Vulcano et al. 2010 is among the first empirical works to study the effectiveness of choice-based revenue management (RM) models. Using data from a major US airline, they report around 5% revenue gains by applying choice-based RM. Using transaction-level data from a major U.S. automaker, Farias et al. (2013) proposes a non-parametric approach that enables data to automatically choose the best choice model for revenue predictions and demonstrates a 20% improvement in prediction accuracy over state-of-the-art benchmark models. Wang (2018) empirically demonstrates that, after incorporating various reference prices into an MNL model, the optimal policies for the assortment planning problems significantly improve the goodness of fit and prediction accuracy of consumer choice behavior. In addition, Feldman et al. (2019) shows that an MNL-based assortment optimization model generates 28% higher revenue per customer visit than the machine-learning-based algorithm currently used by Alibaba's marketplaces. In our paper, we study the optimal layout of photos while allowing for interaction among photos in the layout, i.e., the synergy between cover and non-cover images. We formulate the photo layout optimization problem, which is similar to assortment optimization, as a non-linear integer programming problem. We then provide an algorithm that solves for the optimal solution.

Our paper studies the impact of photo layout on lodging demand on the Airbnb platform and is, therefore, related to the recent literature on property-sharing plat-

forms. In particular, Zhang et al. (2017, 2019) are most relevant to our work. Zhang et al. (2017) employs a quasi-experimental method to examine how the quality of images displayed on Airbnb can increase the property demand by 14.3%. Zhang et al. (2019) shows that, compared to high-quality photos, medium-quality images generate a higher number of reviews and have a more significant effect on property demand in the long run as they are less likely to create a dissatisfactory gap between the perceived and the true quality of the property. Our paper observes that high-quality photos give rise to a higher review writing probability and thus higher demand than medium-quality photos. We arrive at divergent results, possibly because our paper has different data granularity and uses a different empirical strategy than that of Zhang et al. (2019). First, we do not have monthly time stamps on the number of review data. Second, we develop a pairwise comparison model that accounts for changes in the consideration set. Yet, we do not incorporate features such as "effort level" in our model, which may also contribute to the difference in the result. There are also several empirical papers about the Airbnb platform that relate to our work. Farronato and Fradkin (2018) models and estimates the effects of enabling peer supply from Airbnb and shows that the welfare gains are concentrated when hotels have constrained capacities. Li et al. (2016) empirically demonstrates that behavioral differences between nonprofessionals and professional hosts on Airbnb can lead to different revenue. Cui et al. (2019) shows through field experiment that information transparency such as reviews can reduce discrimination by apartment owners on Airbnb. Our work contributes to the existing literature on the property-sharing platform on two fronts. First, we investigate how, in addition to image quality, the image content and the display order posted by Airbnb hosts affect lodging demand. Second, because each listing's providing only one unit of availability per day on the Airbnb platform will bias the estimation results, we develop a novel estimation framework to reduce such estimation bias and consistently estimate the impacts of photo layout and apartment characteristics on the listings' demand.

3.2 Empirical Analysis

3.2.1 Data Description and Empirical Strategy

For our empirical analysis, we use three datasets containing 10,280 Airbnb listings in New York City posted in 2018. The first dataset includes property characteristics, the second contains daily transaction histories for each property, and the third has images of each property.[2]

The first dataset contains information for each individual listing, including listing titles, listing types (entire apartment, private room, or shared room), ZIP codes, overall rating, number of reviews, whether the listing owner is a superhost[3], response rate, number of bedrooms and bathrooms, cancellation policy, etc. A summary of characteristics of the first dataset is presented in Table 3.1.

Table 3.1: Summary of Listings Characteristics

Property characteristics	Mean	St. Dev.	Min	Max
Occupancy Rate	0.52	0.40	0.00	1.00
Price (in USD)	148.38	131.36	10.00	5,068.71
Number of Reviews	28.67	35.97	0	313.00
Overall Rating	4.65	0.33	1.00	5.00
Number of Bedrooms	1.15	0.70	0.00	9.00
Number of Bathrooms	1.12	0.38	0.50	6.50
Response Rate†	93.82	14.69	0.00	100.00
Superhost	0.13	0.34	0.00	1.00
Number of Photos	14.67	10.07	1.00	178.00

† Note: Response rate is defined as the percentage of time a host responds to potential guests within 24 hours. A host becomes as a Superhost if the host satisfies a series of criteria set by Airbnb, such as high overall rating and low cancellation rate.

The second dataset contains daily transaction information for all 10,280 listings in 2018. For each day, a listing may have three possible codes: "A," "B," or "R." "A" indicates that the listing is available on the day, but no customer ends up booking the property. "B" specifies that the listing is blocked and unavailable to customers. "R" refers to a property that is reserved. We can also observe the daily booking

[2]Our datasets are scraped in a socially responsible fashion that will not affect the operational performance of the servers of the Airbnb platform.

[3]A host becomes as a Superhost if the host satisfies a series of criteria set by Airbnb. The detailed criteria are provided at https://www.airbnb.com/superhost.

Table 3.2: Example of Daily Transaction Data

Property ID	Check-in Date	Status	Transaction Date	Price
1000070	2018-09-07	R	2018-05-11	55
1000070	2018-09-08	R	2018-06-11	55
1000070	2018-09-09	A	NA	55
1000070	2018-09-10	B	NA	NA

price and two important dates: (i) *Check-in Date*, which is the date the customer is expected to be accommodated, and (ii) *Transaction Date*, which is defined as the date the customer completes the booking online. The example shown in Table 3.2 shows the status of a listing for four consecutive days. As indicated in the first two rows of Table 3.2, the listing is booked on both September 7th and 8th, yet the booking transactions are completed on May 11th and June 11th, respectively. The third row shows that the property was available on September 9th, indicating that no one booked the property. The listing is blocked and made unavailable for customers on September 10th (row 4 of Table 3.2).

The third dataset includes 222,144 images from all 10,280 listings. For each image, we record the order in which the image is displayed on the webpage. For example, an image with the label "1" indicates that the image is posted as the cover photo, and the label "2" signifies that the image is displayed right after the cover photo. In addition, we use computer vision techniques to identify and evaluate the quality and room type of each image. Image quality is assessed on a 1-7 Likert scale based on how visually pleasant the image is, and room type is classified as *BEDROOM*, *LIVING ROOM*, *OUTSIDE*, *TOILET* or *KITCHEN*. In what follows, we briefly introduce the computer vision models used to recover image characteristics.

3.2.2 Determination of Image Quality and Image Content

We build two separate supervised learning models to efficiently extract image information: one regression model to determine image quality and one classification model to identify room type. Out of the 222,144 images, we randomly selected a subset of 4,000 images, each scored by four subjects (raters) based on how visually pleasing

the picture is.[4] Scores are first pre-processed to exclude outliers, and for each image, the average of the four scores is computed as the image's quality. The labeled score reflects the subject's evaluation of the visual attractiveness of the image and serves as a subjective proxy for photo quality. As a result, the score for the same photo can vary across subjects. We also construct an alternative photo quality measure based on objective photo features such as saturation and hue as a robustness check. The construction of the alternative quality score and the related analysis are presented in Appendix B.2 and Section A of the Online Companion. Subjects are also asked to label the room type based on whether the main part of the image reflects *BEDROOM*, *LIVING ROOM*, *OUTSIDE*, *TOILET* or *KITCHEN* (the photo types are abbreviated as *B*, *L*, *O*, *T*, *K*, respectively). The final room type label is determined based on the majority rule, i.e., the label identified by the majority of the four subjects.[5]

We split the 4,000 images into a training set (60%) and a test set (40%), and we apply convolutional neural networks (CNN), a widely adopted deep learning framework in the computer vision field, to train both models. Since training and tuning a CNN model from scratch requires a great deal of time and effort, we employ transfer learning and a widely used CNN framework named Resnet50. Resnet50 has a relatively small number of layers and parameters yet outperforms other computer vision models and even humans on standard computer vision tasks (He et al. 2016).

To address the potential overfitting issue when training these two models, we perform standard data augmentation processes (which simulates the actual photo-taking process and enlarges the training set), including random rotation of photos within 20 degrees, horizontal photo flips, and random crops. To test out-of-sample performance, we train the parameters of Resnet50 using 60% of the data. We build one regression model to predict images' quality scores and one classification model to predict the room types featured in the images. Table 3.3 presents the summary statistics of the image characteristics.

In Panel (a) of Figure 3-2, we plot the discrepancy between the predicted and ac-

[4]The detailed labeling instruction closely follows the guidelines provided in Zhang et al. (2017), which we attach in the Online Companion.

[5]In the case of a draw, we randomly choose a label between the two labels proposed.

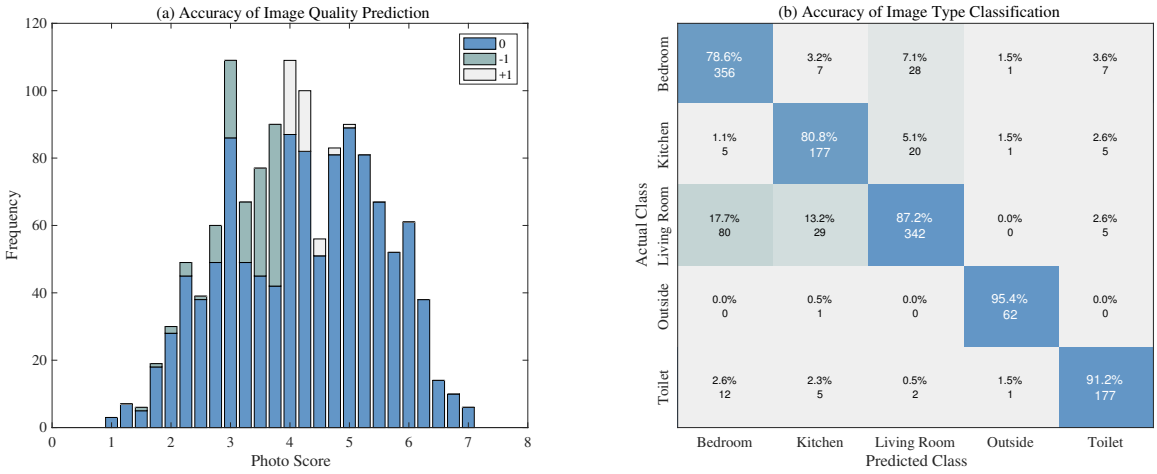

Figure 3-2: Accuracy of Image Quality Identification

tual scores, which takes the value of 0, +1, or −1, where 0 indicates that the predicted score equals the actual one, and +1(−1) signifies that the predicted score is higher (lower) than the actual one. Panel (a) of Figure 3-2 also shows that the misspecified images are concentrated around the threshold, which reflects the subjectivity and ambiguity of image quality judgments. The overall out-of-sample prediction accuracy for photo quality is 85.7% when we split the quality scores into high and low while setting the threshold to 4. Panel (b) of Figure 3-2 shows a confusion matrix of the out-of-sample performance for the image room type classification model. The two numbers in each block represent the percentage and the number of images of the target type classified into the output type. Overall, our photo type classification model achieves an accuracy of 84.0% on the test set. We explain the potential sources of errors in the Appendix B.1. After the training process, we re-train the two models on the subset of 4,000 images and use the resulting models to predict the quality scores and room types for 222,144 images. Table 3.3 summarizes the image quality of each room type. Additionally, Figure 3-3 plots the average image quality score against the image index (the order in the photo is presented on the website). We observe a clear, decreasing pattern of average scores in the display sequence (i.e., cover photos receive the highest quality scores, with decreasing scores associated with each subsequent index). Moreover, the average scores for each listing's first five images are signifi-

Table 3.3: Summary of Image Quality and Content

Statistic	Mean	St. Dev.	Min	Max
Average Image Quality Score	4.069	0.728	1.309	6.429
Number of Living Room Images	5.056	4.133	0.000	69.000
Number of Bedroom Images	3.991	3.436	0.000	66.000
Number of Outside Images	1.820	3.088	0.000	61.000
Number of Toilet Images	1.809	1.569	0.000	22.000
Number of Kitchen Images	1.995	1.788	0.000	45.000

Number of Observations: 10280

cantly higher than the average scores for all images included in each listing. This implies that listing owners are consciously adjusting the sequence of their images so that higher-quality images are prioritized. In addition, the average quality score for the first image is considerably higher, hinting at the significant impact of the cover image. Based on such observation, we formally define the impact of photo layout in the next subsection.

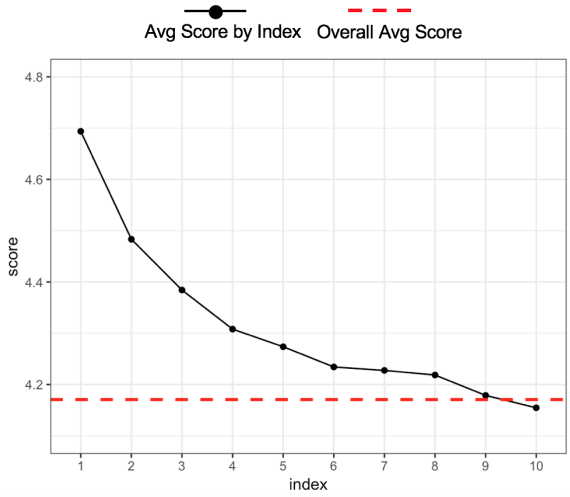

Figure 3-3: Relationship between Average Score and Image Indices

3.2.3 Definition and Identification of Photo Layout

With the photo quality and room types extracted, we quantify the overall impact of the photo layout. Since 2018, listings on the Airbnb platform display the first five pictures of the property as shown in Figure 3-1, where the cover image occupies half of the space, and the other four images evenly split the other half. Given such a

layout, we limit the scope of our analysis to the first five images.[6] Since the cover photo occupies a large area on the web page and is the only image on display when customers browse their search results, we propose separate coefficients indicating if an image of a particular room type is posted in the cover spot or on one of the other four non-cover spots. Specifically, we denote α_m^c, $m \in \{B, L, O, T, K\}$ as the impact of the cover image scores for an image of type m, and α_m^{nc} as the impact of the score for an image of type m when the image is placed in the other four non-cover spots.[7]

When a listing owner posts multiple images of the same room type, however, the value of each additional image could either be discounted as customers may receive less new information from the next image, or strengthened, as customers could confirm the quality of the room after examining it from different angles. Therefore, we introduce δ_m, $m \in \{B, L, O, T, K\}$ to capture the potential diminishing or increasing returns for posting photos of identical room types. Because the prominent position of the cover image receives more attention than photos in the other positions (which makes it more likely to create synergy with the same type of photos at non-cover spots), we assume that the duplicate effect exists when the listing owner posts photos of the same room type in both cover and non-cover slots.

To quantify the overall impact of the photo layout, we incorporate three elements of a photo in our formulation: the quality of the photo, the room type of the photo, and whether or not the photo is posted as the cover image. Specifically, we denote the score for the cover image as s^c, and for ease of notation let $s_m^c = s^c \mathbb{1}_{\{TYPE=m\}}$. Furthermore, we denote the total number of photos of room type m as K_m, where $K_m = K_m^c + K_m^{nc}$, the sum of type-m photos in both cover and non-cover positions. The average and the total score of photos of room type m are $\bar{s}_m$ and S_m where

[6] Seeing more photos requires extra clicking and scrolling on the web page, which dampens the impact from the rest of the photo layout. Meanwhile, including too many images in the layout would weaken the identification.

[7] On Airbnb's mobile application, images are not displayed as according to Figure 3-1. However, customers on the mobile app still observe the cover image for each listing on both the searching page and the property page, and they need to scroll to see the rest photos. Thus, the way we classify layout into cover and non-cover positions still holds for mobile users, whom we cannot distinguish from the web users due to the limitation of our data.

$S_m = \bar{s}_m K_m$. The total impact of the first five photos is then:

$$V = \sum_m \left(\alpha_m^c s_m^c + \alpha_m^{nc} S_m^{nc} + \delta_m S_m^{nc} \mathbb{1}_{\{K_m^c = 1\}} \mathbb{1}_{\{K_m^{nc} \geq 1\}} \right) . \tag{3.1}$$

The first term in Equation (3.1) captures the impact of cover image of room type m and score s_m^c. The second term reflects the total impact of type-m photos displayed at non-cover positions. To better understand the third term, we rewrite it as $\delta_m \bar{s}_m^{nc} K_m^{nc} \mathbb{1}_{\{K_m^c = 1\}} \mathbb{1}_{\{K_m^{nc} \geq 1\}}$, which is non-zero only when the cover image shares the same room type as one or more photos in the non-cover spots. Thus, the third term captures the impact of photo repetition, with the effect size proportional to the number of repeated photos with the same room type and their average photo quality.[8]

Identification of Parameters We now show that the parameters in Equation (3.1) are identifiable. The identification of α_m^c is achieved by the variation in the quality of each room type cover image. Similarly, α_m^{nc} is identifiable because of the variation in the number of photos and the photo quality in non-cover spots. Finally, the identification of δ_m is achieved through the variation of the room type in the cover image position, the number of photos that have the same room type, and their quality scores.

3.2.4 Reduced-form Analysis

We now conduct reduced-form regression analysis to study the impact of photo layout on property occupancy rate. From the dataset, we first construct the occupancy rate, denoted as OC_RATE, to reflect the popularity of a listing:

$$OC\_RATE_{it} = \frac{NR_{it}}{NR_{it} + NA_{it}}, \tag{3.2}$$

where NR_{it} is the number of days property i is reserved in month t and NA_{it} the number of days the property is available in month t. In our reduced-form regressions,

[8] We assume the impact of the cover image will not be affected in this case, so we use the average score for only the non-cover images.

Table 3.4: Regression Results: Impact of Image Layout on Occupancy Rates

	(1) OLS			(2) 2SLS		
Photo Level	**Cover**	**Non-Cover**	**Duplicate**	**Cover**	**Non-Cover**	**Duplicate**
Bedroom	1.10***	0.29***	-0.006	0.99***	0.24***	0.002
	(0.17)	(0.05)	(0.05)	(0.17)	(0.05)	(0.05)
Living_Room	0.94***	0.11*	0.11*	0.89***	0.09	0.11*
	(0.16)	(0.05)	(0.05)	(0.16)	(0.05)	(0.05)
Outside	0.87***	-0.13	0.38***	0.87***	-0.15*	0.38***
	(0.20)	(0.07)	(0.11)	(0.20)	(0.07)	(0.11)
Toilet	0.90*	0.22***	-0.07	0.72	0.21**	-0.01
	(0.46)	(0.07)	(0.47)	(0.46)	(0.07)	(0.47)
Kitchen	0.94***	0.27***	0.32**	0.88***	0.23***	0.32*
	(0.20)	(0.06)	(0.12)	(0.20)	(0.05)	(0.12)
Property Level						
Price ($\times 10^{-2}$)	-1.50***			-1.80***		
	(0.05)			(0.15)		
Overall_Rating	5.68***			6.33***		
	(0.42)			(0.45)		
Num_of_Rev	0.19***			0.21***		
	(1.74×10^{-3})			(7.53×10^{-3})		
Superhost	5.41***			4.52***		
	(0.27)			(0.36)		
Response_Rate	0.24***			0.22***		
	(0.06)			(0.07)		
Num_of_Photos	-0.11***			-0.13***		
	(0.01)			(0.01)		
Num_of_Bedrooms	1.70***			1.99***		
	(0.17)			(0.19)		
Num_of_Bathrooms	0.13			0.41		
	(0.29)			(0.30)		
IV	NO			YES		
Monthly FE	YES			YES		
Zipcode FE	YES			YES		
Observations	86,142			86,142		
Adjusted R^2	0.296			0.294		
Note:						*p<0.1; **p<0.05; ***p<0.01

we assume the monthly occupancy rate is linear in the characteristics; in other words:

$$OC\_RATE_{it} = \beta_p p_{it} + V_i + \boldsymbol{\tau}\ PROPERTY_i + \boldsymbol{\gamma}\ CONTROL_{it} + \epsilon_{it}\,, \tag{3.3}$$

where p_{it} is the average price in month t for the listing i. V_i is the impact of photo layout defined in the previous section, where

$$V_i = \sum_m \left(\alpha_m^c s_{mi}^c + \alpha_m^{nc} S_{mi}^{nc} + \delta_m S_{mi}^{nc} \mathbb{1}_{\{K_{mi}^c=1\}} \mathbb{1}_{\{K_{mi}^{nc}\geq 1\}}\right) \tag{3.4}$$

. We assume that photo layout for each listing remains the same across different time periods. *PROPERTY* is a vector of property characteristics, including overall rating by Airbnb guests, number of reviews, whether the owner is a superhost, response

rate, number of bedrooms and bathrooms and the total number of photos each listing owner posts. $\boldsymbol{\tau}$ depends only on the property and is time invariant, and *CONTROL* includes monthly fixed effects and location (ZIP code) fixed effects.

The regression results are presented in column (1) of Table 3.4. Lower prices, higher image quality, a higher number of bedrooms, and higher ratings by Airbnb guests are associated with higher monthly occupancy rates. Also, listings owned by superhosts and those with higher response rates are more likely to be booked. Most importantly for this study, cover images, in general, have a much higher impact on occupancy rates than images in non-cover spots, likely because cover images are displayed on the search results page and take up more space on the listing page than the other four photos. More specifically, high-quality bedroom cover images lead to the biggest increase in occupancy rates among the five room types. Finally, the positive coefficients of the duplicate effect from the living room, outside, and kitchen images indicate that having identical photos can lead to greater improvements in demand than the combined improvements from individual photos.

Endogeneity In our regression analysis, both the rental price of the property and the number of reviews suffer from potential endogeneity concerns as there could be unobserved property characteristics that affect listing owners' pricing decisions and customers' reviews. To address the endogeneity in price, we use as an instrumental variable (IV) the average price of competing listings that share the same property type (such as apartment or loft) and are located in the same neighborhood. This IV is correlated with the price of the focal listing as the competing listings that have the same property type and are located in the neighborhood charge similar prices as the focal listing. In addition, the average price of competing listings is unlikely to be correlated with the unobserved characteristics of the focal apartment. To address the endogeneity in the number of reviews, we use the average number of reviews of competing listings as the IV. According to Table 3.4, the main results remain unchanged after we incorporate these two IVs in our two-stage least squares (2SLS) regression. Meanwhile, both the 1st stage F-statistic and the Cragg-Donald Statistic presented in Table 3.5 reject the null hypothesis, suggesting that the instrument is

Table 3.5: Endogenous Variables, Instrument Variables and Related Tests

Endogenous Variable	IV	1st Stage F-Statistic	Cragg-Donald Statistic	Wu-Hausman Statistic
Price	Avg. Price‡	6945.95***	2434.10***	11.35***
Number of Reviews	Avg. # Rev‡	2508.60***		

‡The average price and number of reviews are taken over listings of the same type in the neighborhood
Note:*p<0.1; **p<0.05; ***p<0.01

not weak. Furthermore, the Wu-Hausman test is also significant, which implies the consistency of the two IVs.

3.3 Structural Estimation

Our reduced-form regression results demonstrate the impact of photo layout on occupancy rate. This section performs structural estimation to gauge the effect of photo layout on customers' booking decisions to closely replicate customers' booking processes at Airbnb, capture the substitution pattern among listings, and eventually suggest optimal photo layouts.

3.3.1 Identification Strategy

On the Airbnb platform, each property provides only one unit of supply each day. Thus the demand for each listing on each day is censored to 1. Such heavily censored demand data imposes two challenges to our estimation process. First, customers face a changing choice set: when a customer books a property for specific check-in date, the property becomes unavailable. Consequently, the size of choice set for customers who arrive after this transaction will shrink by one. To not bias the estimation, we need to keep track of the availability of each property and incorporate the changing choice set into our estimation model whenever a customer books a property. Second, the demand for each property is heavily censored in our data. Because each property can only be booked once on a specific day, the daily demand is truncated at 1 and cannot fully capture the attractiveness of each listing. To elaborate, consider a case

in which two properties both get booked out every single day during a period of 30 days. According to the booking data, both apartments will be considered equally popular as both listings have 30 units of demand over the period. However, if, in fact, one apartment regularly gets booked ahead of the other, we need to make sure that our estimation model properly reflects that one apartment is preferred over the other.

Traditional estimation approaches, such as the MNL choice model that has a fixed consideration set (McFadden 1978) and the maximum score estimation approach (Fox 2007), fail to converge in this setting due to the challenges mentioned above. To obtain consistent estimates of the impact of image layout, we develop a pairwise comparison model (PCM) - an estimation framework that consistently estimates the impact of property characteristics on demand. In what follows, we present our estimation model, and we demonstrate its asymptotic properties in Section 3.3.1. We also compare the performance of different estimation approaches for the Airbnb platform using synthetic data in Section 3.3.1.

To start with, we assume that the customer k's utility from booking apartment i is

$$\begin{aligned} \mu_{ik} &= \boldsymbol{X_i}\boldsymbol{\beta} + \epsilon_{ik} \\ &= \beta_p p_i + V_i + \boldsymbol{\tau} PROPERTY_i + \boldsymbol{\gamma} CONTROL_i + \epsilon_{ik}\,, \end{aligned} \tag{3.5}$$

for which the utility function is linear in the apartment specific covariates. We assume that ϵ_{ik} are independent and identically distributed (i.i.d.) across listings and customers, yet we do not specify a particular functional form for the error term ϵ_{ik}.

Next, we define $\mathbb{1}_{\{i \succ j\}d}$ as whether property i is reserved ahead of property j on check-in date d according to the data, given both listings are not blocked on that day. For example, suppose we observe both properties i and j are reserved on September 1st in the data. For property i, the booking transaction was completed on August 15th, and for property j, it was August 18th. Then, we have $\mathbb{1}_{\{i \succ j\}d} = 1$ and $\mathbb{1}_{\{j \succ i\}d} = 0$ on September 1st. Alternatively, suppose property i was booked on August 15th, yet no

one ends up booking property j. In this case, we still have $\mathbb{1}_{\{i\succ j\}d} = 1$ and $\mathbb{1}_{\{j\succ i\}d} = 0$ on September 1st. Accordingly, $\sum_{d=1}^{D} \mathbb{1}_{\{i\succ j\}d}$ would give the number of observations where listing i is selected ahead of j over the period of D as recorded in the data, and $(\sum_{d=1}^{D} \mathbb{1}_{\{i\succ j\}d} + \sum_{d=1}^{D} \mathbb{1}_{\{j\succ i\}d})$ is the total number of possible comparisons in the time horizon of D days from the data. For each pair of listings (i, j), given the covariates X_i, X_j, it holds that, in expectation, the theoretically predicted number of times i is selected ahead of j should equal to the observed number of times i is selected ahead of j. That is to say:

$$\begin{aligned}
\mathbb{E}\big[\mathbb{1}_{\{i\succ j\}d}|\boldsymbol{X_i}, \boldsymbol{X_j}\big] &= \mathbb{P}(i \succ j|\boldsymbol{\beta}, \boldsymbol{X_i}, \boldsymbol{X_j})\bigg(\sum_{d=1}^{D} \mathbb{1}_{\{i\succ j\}d} + \sum_{d=1}^{D} \mathbb{1}_{\{j\succ i\}d}\bigg) \qquad (3.6)\\
&= \mathbb{P}((\boldsymbol{X_i} - \boldsymbol{X_j})\boldsymbol{\beta} \geq \epsilon_{jk'} - \epsilon_{ik})\bigg(\sum_{d=1}^{D} \mathbb{1}_{\{i\succ j\}d} + \sum_{d=1}^{D} \mathbb{1}_{\{j\succ i\}d}\bigg)\\
&= G(\boldsymbol{\beta}(\boldsymbol{X_i} - \boldsymbol{X_j}))\bigg(\sum_{d=1}^{D} \mathbb{1}_{\{i\succ j\}d} + \sum_{d=1}^{D} \mathbb{1}_{\{j\succ i\}d}\bigg),
\end{aligned}$$

where $\mathbb{P}(i \succ j|\boldsymbol{\beta}, \boldsymbol{X_i}, \boldsymbol{X_j})$ is the theoretical probability that, given $\boldsymbol{\beta}$, $\boldsymbol{X_i}$ and $\boldsymbol{X_j}$, the property i is booked ahead of j when both properties are made available, i.e., not blocked by the owners. In addition, the link function $G(\cdot)$ is the cumulative distribution function (c.d.f.) of the difference of the error term $\epsilon_{jk'} - \epsilon_{ik}$. For example, when ϵ_{ij} follows Gumbel distribution, $G(\cdot)$ is simply the c.d.f. of the logistic distribution.

For the general case where $G(\cdot)$ may not have a closed-form expression, we estimate Equation (3.6) by nonlinear least squares (NLS):

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \Bigg(\underbrace{\mathbb{P}(i \succ j|\boldsymbol{\beta})\bigg(\sum_{d=1}^{D} \mathbb{1}_{\{i\succ j\}d} + \sum_{d=1}^{D} \mathbb{1}_{\{j\succ i\}d}\bigg)}_{\text{expected number of times } i \text{ is selected ahead of } j} - \underbrace{\sum_{d=1}^{D} \mathbb{1}_{\{i\succ j\}d}}_{\substack{\text{actual number of times } i \text{ is}\\ \text{selected ahead of } j}} \Bigg)^2. \qquad (3.7)$$

We evaluate $\boldsymbol{\beta}$ in Equation (3.7) by minimizing the squared discrepancy between the expected frequency and the realized frequency in the data, over all $n(n-1)/2$ pairs of listings.

PCM overcomes our two estimation challenges of changing consideration sets and

censored demand. Unlike MNL, which requires a fully specified consideration set to specify the probability of a particular option being chosen, the PCM approach needs only a function form for the error term to specify the probability that property i is chosen ahead of property j, which, together with the number of comparisons between property i and j, allow for the identification of the coefficients for the listing features. Additionally, since the demand is censored to be 0 or 1 in the Airbnb data, it is possible that two popular listings both get reserved every day and are considered equal by traditional choice models. In this situation, PCM can identify the more favored listing and the impact of the corresponding features by utilizing the information on the booking sequence on top of quantity information to identify consumers' preferences. Apart from the two abovementioned advantages, PCM is more robust to error term misspecification than other commonly used estimation approaches, which we discuss formally in Section 3.3.1.

Properties of the PCM Estimator

With the estimation framework specified, we now develop two asymptotic properties of the PCM estimator to enable further statistical inferences. We first list three assumptions relevant to our propositions.

Assumption 3.1. *For a nonlinear regression model $y_t = f_t(\boldsymbol{X}, \boldsymbol{\beta}) + \epsilon_t$, assume the following:*
(i) $f_t(\boldsymbol{X}, \boldsymbol{\beta}) = f(\boldsymbol{X}, \boldsymbol{\beta})$, $\frac{\partial f}{\partial \boldsymbol{\beta}}$ exists and is continuous.
(ii) $\boldsymbol{\beta}$ belongs to a compact space.
(iii) $\mathbb{E}[\frac{\partial f}{\partial \boldsymbol{\beta}} \frac{\partial f}{\partial \boldsymbol{\beta}}']$ is finite and nonzero.

All three assumptions hold in our context for the following reasons. In this nonlinear regression framework, each observation corresponds to a pair a listings (i, j), and $\boldsymbol{X}$ represents $(\boldsymbol{X_i} - \boldsymbol{X_j})$, the difference between feature vectors of listings i and j. The dependent variable y_t is the number of observations in the data where property i is booked ahead of j, while $f_t(\boldsymbol{X}, \boldsymbol{\beta})$ is the expected number of observations where property i is booked ahead of j. As Equation (3.7) suggests, this term equals the

multiplication of $\mathbb{P}(i \succ j|\boldsymbol{\beta})$, i.e., the theoretical probability that i is preferred to j given $\boldsymbol{\beta}$, and the total number of comparisons between i and j.

The first part of assumption (i) holds if $f_t(\boldsymbol{X}, \boldsymbol{\beta})$ is time-invariant because in our formulation, the probability function $\mathbb{P}(i \succ j|\boldsymbol{\beta})$ has the same functional form across all the observation pairs. For the second part of assumption (i), we can obtain the derivative of the link function $G(\cdot)$ in Equation (3.6) with respect to $\boldsymbol{\beta}$, which is proportional to the probability distribution function (p.d.f.) of the underlying link distribution. The p.d.f. of $G(\cdot)$ is continuous for most, if not all, error distributions. For instance, when the link function follows a normal distribution, the derivative for f is $\frac{\partial f}{\partial \boldsymbol{\beta}} = (\boldsymbol{X_i} - \boldsymbol{X_j})\phi(\boldsymbol{\beta}(\boldsymbol{X_i} - \boldsymbol{X_j}))$, where ϕ is the p.d.f. of the normal distribution. Assumption (ii) naturally holds, since we expect the true values of $\boldsymbol{\beta}$ to be finite and thus bounded within a compact set. Assumption (iii) assumes that the second moment of $\frac{\partial f}{\partial \boldsymbol{\beta}}$ is non-zero and finite. In our context, this is also valid for most of the link functions. Again, we use the normally distributed link function as an example. $\frac{\partial f}{\partial \boldsymbol{\beta}} = (\boldsymbol{X_i} - \boldsymbol{X_j})\phi(\boldsymbol{\beta}(\boldsymbol{X_i} - \boldsymbol{X_j}))$ is the multiplication of the difference in feature vectors and the p.d.f.. As the feature difference $(\boldsymbol{X_i} - \boldsymbol{X_j})$ between any two pairs of listings are bounded, the second moment of this term is non-zero and finite. Thus, assumption (iii) also holds for the commonly used link functions. With the three assumptions justified, we formally state the consistency of the nonlinear least square PCM estimator $\hat{\beta}$ in Proposition 3.1.

Proposition 3.1. *Under Assumption (i) and (ii), the PCM nonlinear least square estimator $\hat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta_0}$, the true parameter in the data generating process.*

Similar to other M-estimators, Proposition 3.1 indicates that the PCM nonlinear least square estimator is consistent. A formal proof of the consistency is presented in Appendix B.4. In addition, we also establish the asymptotic normality of the estimator as the following.

Proposition 3.2. *Under assumptions (i)-(iii), the PCM least square estimator $\hat{\boldsymbol{\beta}}$ is*

asymptotically normal with

$$\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta_0}) \xrightarrow{d} \mathcal{N}(0, \sigma_0^2 \mathbb{E}[\frac{\partial f}{\partial \boldsymbol{\beta}} \frac{\partial f}{\partial \boldsymbol{\beta}}']^{-1}) \,, \tag{3.8}$$

where σ_0 is the standard deviation of the discrepancy between the theoretical and the observed probabilities in Equation (3.7).

Again, the proof of the proposition is presented in Appendix B.4. The asymptotic normality enables the construction of confidence intervals and justifies our statistical inference. Notice that the asymptotic property holds for the normally distributed utility error term and for a family of distributions that satisfy assumptions (i)-(iii). However, to interpret the convergence rate of the estimator, we assume that the error term follows a normal distribution for the present. Recall that $\mathbb{E}[\frac{\partial f}{\partial \boldsymbol{\beta}} \frac{\partial f}{\partial \boldsymbol{\beta}}']$ is the second moment of $(\boldsymbol{X_i} - \boldsymbol{X_j})\phi(\boldsymbol{\beta}(\boldsymbol{X_i} - \boldsymbol{X_j}))$. When the variation in listing features $(\boldsymbol{X_i} - \boldsymbol{X_j})$ is very small, the convergence rate is slow as the identification of parameters is weak. If there's no variation in the features, assumption (iii) will be violated, and the model will fail to converge at all. On the other hand, when $(\boldsymbol{X_i} - \boldsymbol{X_j})$ is very large, the term $\phi(\boldsymbol{\beta}(\boldsymbol{X_i} - \boldsymbol{X_j}))$ becomes small, and, consequently, the convergence slows down. The underlying reason for this is that when the listing features are distributed very sparsely in the feature space, there will be a range of $\boldsymbol{\beta}$ that would yield the same observations. More data points are required to recover the true parameter $\boldsymbol{\beta}$.

Special Case: Gumbel or Normal Error Term.

We now discuss two special cases in which the error term ϵ_{ik} in Equation (3.5) follows $Gumbel(0, 1)$ and $\mathcal{N}(0, 1)$, respectively. Specifically, we show how these two commonly used distributions lead to simplified estimation process. We start with the case where $\epsilon_{ik} \sim Gumbel(0, 1)$. In this case, after algebaric manipulation, Equation (3.6) becomes:

$$\mathbb{E}\left[\frac{\sum_{d=1}^{D} \mathbb{1}_{\{i \succ j\}d}}{\sum_{d=1}^{D} \mathbb{1}_{\{i \succ j\}d} + \sum_{d=1}^{D} \mathbb{1}_{\{j \succ i\}d}} \middle| \boldsymbol{X_i}, \boldsymbol{X_j}\right] \;=\; G(\boldsymbol{\beta}(\boldsymbol{X_i} - \boldsymbol{X_j})) \tag{3.9}$$

$$
\begin{aligned}
&= \mathbb{P}((\boldsymbol{X_i} - \boldsymbol{X_j})\boldsymbol{\beta} \geq \epsilon_{jk'} - \epsilon_{ik}) \\
&= \frac{1}{1 + \exp\left(-(\boldsymbol{X_i} - \boldsymbol{X_j})\boldsymbol{\beta}\right)},
\end{aligned}
$$

where k and k' are the indices for customers who reserve properties i and j. Because both ϵ_{ik} and $\epsilon_{jk'}$ follow i.i.d. $Gumbel(0,1)$, the difference between the two error terms, $\epsilon_{jk'} - \epsilon_{ik}$, follows $Logistic(0,1)$. This way, Equation (3.9) can be directly estimated through fractional response logistic regression (Papke and Wooldridge 1996) with common built-in functions such as **glm()** in R.

Similarly, when $\epsilon_{ik} \sim Normal(0,1)$, Equation (3.6) can be written as:

$$
\mathbb{E}\left[\frac{\sum_{d=1}^{D} \mathbb{1}_{\{i \succ j\}d}}{\sum_{d=1}^{D} \mathbb{1}_{\{i \succ j\}d} + \sum_{d=1}^{D} \mathbb{1}_{\{j \succ i\}d}} \middle| \boldsymbol{X_i}, \boldsymbol{X_j}\right] = G(\boldsymbol{\beta}(\boldsymbol{X_i} - \boldsymbol{X_j})) = \Phi(\boldsymbol{\beta}(\boldsymbol{X_i} - \boldsymbol{X_j})) \quad (3.10)
$$

where $\Phi(\cdot)$ is the c.d.f. of normal distribution. As both ϵ_{ik} and $\epsilon_{jk'}$ follow i.i.d. $\mathcal{N}(0,1)$, and the difference of $\epsilon_{jk'} - \epsilon_{ik}$ follows $\mathcal{N}(0, \sqrt{2})$. As a result, Equation (3.10) can then be estimated by fractional response probit regression.

Performance Comparisons for Estimation Models

Before presenting our estimation results, we first demonstrate the performance of the PCM estimator. To this end, we benchmark PCM with three different estimation models on synthetic data under two different specifications in the Airbnb setting. The data generating process follows the random utility model proposed by Fox (2007):

$$
\mu_{ai} = -x_{1,ai} + \beta_0 x_{2,ai} + \tau_{ai}\,, \quad (3.11)
$$

where μ_{ai} is the utility gained for customer a choosing listing i, $x_{1,ai}$ and $x_{2,ai}$ are the covariates and are i.i.d. normally distributed as $\mathcal{N}(0,2)$, β_0 is the parameter to be recovered by the estimation models and has true value $\beta_0 = 1$, and τ_{ai} is the error term. In the first specification, τ_{ai} follows a standard Gumbel distribution. In the second specification, τ_{ai} follows a bimodal distribution with the same mean and variance as the standard Gumbel distribution. Similar to Section 3.3.1, after the

utility for each property is realized, customers arrive sequentially to decide which property to book. When a property is booked, the consideration set shrinks by one, accordingly.

Multinomial Logit Model (MNL) We use the widely adopted MNL model as a benchmark. Under the MNL model, the probability of choosing listing i is $P_i = \exp(\boldsymbol{\beta X_i})/\sum_{i'}\exp(\boldsymbol{\beta X_i'})$.

Sequential Multinomial Logit Model (SMNL) One of the limitations of the classic MNL model is the fixed consideration set that the model assumes. When applied to the Airbnb setting, the changing consideration sets caused by properties having only unit availability will lead to biased estimation results. To address this issue, we modify the classic MNL to take into account the changing choice sets, and we propose a sequential MNL (SMNL) approach. The probability of choosing i on day t is:

$$\begin{aligned} P_i^d = \mathbb{P}(U_i \geq U_{i'}, \forall i' \in \mathbb{I}^d) &= \frac{\exp(U_i)}{\sum_{i'\in\mathbb{I}^d}\exp(U_{i'})} = \frac{\exp(\boldsymbol{\beta X_i})}{\sum_{i'}\exp(\boldsymbol{\beta X_i'})\mathbb{1}_{B_i^d \leq B_{i'}^d}} \quad (3.12) \\ \hat{\boldsymbol{\beta}} &= \arg\max_{\boldsymbol{\beta}} \sum_t \sum_i \log P_i^d, \end{aligned}$$

where U_i is the mean utility, and $\mathbb{I}^d$ is the choice set when i is booked on day d, and it shrinks over time. This shrinking behaviour is captured by $\mathbb{1}_{B_i^d \leq B_{i'}^d}$, which is an indicator of whether listing i gets booked ahead of i' on the check-in date d. Then we can use maximum likelihood estimator (MLE) to find the optimal $\hat{\beta}$. Note that after we introduce the indicator, the computational time becomes much longer for high dimensional $\hat{\boldsymbol{\beta}}$.

Maximum Score (MS) Another limitation of the MNL model is its assumption that the error term follows a Gumbel distribution. To mitigate potential mis-specification of the error term, Fox (2007) proposes a semi-parametric maximum score (MS) model:

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \sum_{i=1}^{K-1}\sum_{j=i+1}^{K}\frac{1}{N}\sum_{a=1}^{N}\left(\mathbb{1}_{[ai]}\mathbb{1}_{[\boldsymbol{\beta X_{ai}} > \boldsymbol{\beta X_{aj}}]} + \mathbb{1}_{[aj]}\mathbb{1}_{[\boldsymbol{\beta X_{aj}} > \boldsymbol{\beta X_{ai}}]}\right), \quad (3.13)$$

where the expression enumerates all pairs of listings (i, j). $\mathbb{1}_{[ai]}$ indicates whether list-

ing i is chosen at observation a, and $\hat{\boldsymbol{\beta}}$ forms a deterministic ranked preference list for the expression to be most consistent with data. Fox (2007) shows that the MS estimator is asymptotically consistent and very fast to compute. However, identification of the true parameter with limited data is challenging.

Using synthetic data where listings have unit availability every day, we run the three abovementioned estimation approaches together with our PCM model (estimated both through NLS and GLM). Adopting the data setup in Fox (2007), we generate 100 listings with normally distributed features, and the parameter to be estimated has true value $\beta_0 = 1$. We assume that there are 500 potential customers. Upon each customer's arrival, an i.i.d. error term is realized for each listing, and the listing with the highest realized utility will be booked. If the realized utility for all available listings is lower than that of the outside option (with mean utility normalized to zero), that demand is lost. Once a listing is booked on a day, it will be unavailable for subsequent customers. Meanwhile, we consider two error term distributions for listings' utility. In the first case, the error term is a standard Gumbel distribution. In the second case, the error term is misspecified as standard Gumbel but is, in fact, a mixed-normal distribution with the same mean and variance as the standard Gumbel distribution. To reflect the scope of the Airbnb dataset and our estimation process, in the synthetic data analysis, we account for the outside option in the data-generating process but not in the estimation stage. Customers compare different options and choose the one with the highest realized utility, be it a listing on Airbnb or the outside option. Then, we perform conditional PCM estimation using only the simulated booking history on Airbnb. We report the estimation results in Table 3.6.

The estimation results show that the classic MNL model fails to recover the true value of β_0, as the model does not account for the changing consideration sets. The MS estimator performs poorly due to the lack of data and failure to incorporate consideration set shrinkage. When the error term is correctly specified, both the SMNL and PCM estimators can recover a parameter value close to $\beta_0 = 1$, with PCM presenting a smaller gap between the recovered parameter and the true value of

Table 3.6: Estimation Results Using Synthetic Data ($\beta_0 = 1$, 95% CI)

Error Structure	Gumbel			Mixed-Normal		
Choice Model	Mean	St. Dev.	RMSE	Mean	St. Dev.	RMSE
MNL	0.095	0.194	0.925	0.147	0.213	0.878
SMNL	0.995	0.016	0.017	1.206	0.049	0.211
MS	0.170	3.463	3.544	0.885	3.151	3.137
PCM (NLS)	0.997	0.014	0.015	1.073	0.022	0.076
PCM (GLM)	1.003	0.021	0.021	1.171	0.045	0.177

β_0. However, when the error structure is misspecified, PCM (NLS) appears to be the most robust approach, as it presents the smallest gap with the true value. Finally, when estimated through GLM, the running time of PCM is significantly reduced, yet a larger bias is reported when the error term is misspecified. In Appendix B.3, we present additional numerical experiments to showcase that PCM can uncover the ground-truth parameter regardless of whether outside option observations are included.

3.3.2 Structural Estimation of Parameters and Estimation Results

Next, we conduct structural estimation by applying PCM to the transaction data from all ten neighborhoods in Manhattan. The 10 neighborhoods shown in Figure 3-4 encompass 43 ZIP codes that include 2,561 Airbnb listings and 297,815 daily transactions.[9] We assume that customers compare lodging alternatives within each neighborhood, i.e., listings in areas that have the same color code in Figure 3-4 are compared pair-wised. In addition, we assume that customers' preferences for listings characteristics over the ten neighborhoods are homogeneous, i.e., they share the same set of parameters in our estimation. The estimates we obtain represent the average effects of photos and other characteristics on the listing attractiveness. While different regions can have heterogeneous fixed effects, those fixed effects will get canceled in the estimation process because we perform pairwise comparisons on listing pairs within

[9]Many listings labeled as a *private room* do not display images of a kitchen or living room. Therefore, to examine the impact of all five room types, our study focuses only on listings that rent the entire home/apartment.

each region. Furthermore, as an existing listing may switch its availability status, the consideration set may not be strictly shrinking over time. Thus, we only focus on the demand within one month before the check-in date in our estimation, as listing owners are less likely to switch the status of their listing as the check-in date approaches. We provide a detailed discussion on this issue in Appendix B.2.

To exploit the flexibility of the error term structure in Equation (3.5), we test two specifications where the error term in Equation (3.5) follows Gumbel and normal distribution, respectively. We estimate $\boldsymbol{\beta} = (\beta_p, \boldsymbol{\tau}, \boldsymbol{\gamma}, \alpha_m^c, \alpha_m^{nc}, \delta_m)$ by combining the customer utility specified in Equation (3.5) with the pairwise estimation Equation (3.7). Specifically, our estimation equation is as follows:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{m=1}^{12} \sum_{\eta=1}^{10} \sum_{(i,j)\in S(m,\eta)} \left(\mathbb{P}(i \succ j|\boldsymbol{\beta}) \left(\sum_{d=1}^{D} \mathbb{1}_{\{i\succ j\}d} + \sum_{d=1}^{D} \mathbb{1}_{\{j\succ i\}d} \right) - \sum_{d=1}^{D} \mathbb{1}_{\{i\succ j\}d} \right)^2, \tag{3.14}$$

where we sum over 12 months, 10 neighborhoods, and all the listing pairs in month m and neighborhood η, respectively. Additionally, $\mathbb{P}(i \succ j|\boldsymbol{\beta})$ follows logistic (normal) distribution when the error term in Equation (3.5) follows Gumbel (normal) distribution. We further use bootstrap to obtain 95% confidence intervals for our estimates.

We present the estimation results in Table 3.7. As the main insights from both specifications are identical, yet the model under the Gumbel error structure presents a higher pseudo-R-squared value, we will focus on the results under the Gumbel error term in what follows. The results suggest that the bedroom room type is the best choice for the cover image, given the same photo quality. When the quality of the bedroom cover image increases by 1, the customer's utility increases by 0.108 - the highest magnitude among all room types. This is likely because most Airbnb customers stay overnight and place great value on the quality and coziness of the bedroom. Therefore, for most customers, a high-quality picture of the bedroom will likely be particularly attractive and play a significant role in their decision-making processes. Kitchen and outside views are the second and third best options, respec-

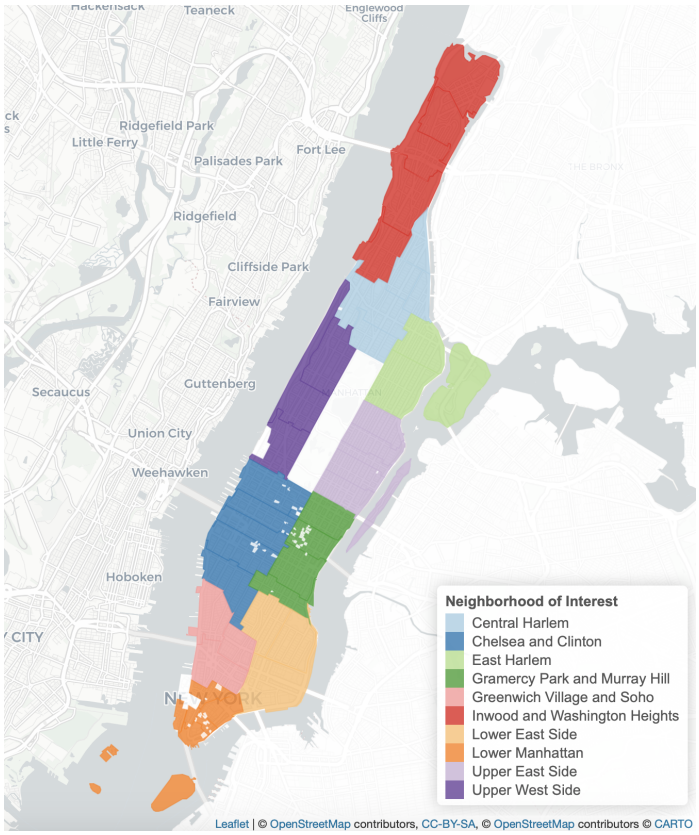

Figure 3-4: Neighborhoods of Interest in Manhattan.

tively, for the cover image. Interestingly, posting living room photos as cover images results in the lowest increases in demand, even though more than 60% of the listings currently use living room cover images. There were 12 incidences in which listing owners posted a toilet as the cover image. We removed all 12 incidences from our estimation due to weak identification for the toilet cover photo.

The magnitudes of coefficients of the non-cover images are less than half than their cover counterparts, highlighting the crucial role that the cover images play in helping customers decide which property to book. As noted above, these cover images are displayed on the search results webpage to attract customers to click on the listing for further inspection, and they take up more space on the listing webpage than the additional four photos. Bedroom (living room) photos have negative (positive) duplicate effect coefficients, implying decreasing (increasing) marginal returns when posting photos of these room types in both cover and non-cover spots. The positive

Table 3.7: Structural Estimation Results

	(1) Gumbel Error Terms			(2) Normal Error Terms		
Photo Level	**Cover**	**Non-Cover**	**Duplicate**	**Cover**	**Non-Cover**	**Duplicate**
Bedroom	0.108	0.028	-0.009	0.086	0.024	-0.006
	(0.100 0.119)	(0.025 0.030)	(-0.012 -0.006)	(0.056 0.098)	(0.022 0.026)	(-0.009 0.000)
Living_Room	0.077	0.022	0.006	0.063	0.020	0.005
	(0.070 0.085)	(0.019 0.026)	(0.004 0.009)	(0.046 0.072)	(0.018 0.023)	(0.002 0.007)
Outside	0.084	0.014	NA	0.071	0.013	NA
	(0.074 0.097)	(0.008 0.020)	(NA)	(0.062 0.079)	(0.009 0.017)	(NA)
Toilet	NA	0.017	NA	NA	0.014	NA
	(NA)	(0.009 0.022)	(NA)	(NA)	(0.010 0.018)	(NA)
Kitchen	0.089	0.026	NA	0.072	0.023	NA
	(0.076 0.102)	(0.023 0.029)	(NA)	(0.055 0.083)	(0.021 0.026)	(NA)
Property Level						
Price ($\times 10^{-2}$)	-0.403			-0.343		
	(-0.415 -0.393)			(-0.362 -0.332)		
Overall_Rating	0.306			0.253		
	(0.293 0.318)			(0.219 0.275)		
Num_of_Rev	0.728			0.624		
($\times 10^{-2}$)	(0.754 0.843)			(0.617 0.631)		
Superhost	0.186			0.151		
	(0.171 0.201)			(0.121 0.172)		
Response_Rate	0.795			0.686		
($\times 10^{-2}$)	(0.754 0.843)			(0.647 0.724)		
Num_of_Photos	0.171			0.151		
($\times 10^{-2}$)	(0.105 0.230)			(0.109 0.212)		
Num_of_Bedrooms	0.227			0.198		
	(0.215 0.239)			(0.184 0.223)		
Num_of_Bathrooms	0.106			0.086		
	(0.095 0.118)			(0.059 0.104)		
Pseudo R Squared:	0.357			0.349		

duplicate effect for living room might be one of the reasons why, albeit having the lowest impact among all room types in the cover position, more than 60% of the listings in our data use living room photos as cover images. We do not analyze duplicate effects for toilet, kitchen and outside pictures due to insufficient number of observations.

Moreover, lower prices, higher overall ratings, and a more significant number of reviews all make properties more likely to be booked. Additionally, listings for which the owners are superhosts and with high response rates are more likely to be booked by the customers. Finally, listings with more bedrooms and bathrooms also enjoy higher demand.

Robustness Check. We perform structural estimation over additional specifications to consolidate our estimation results. The results are summarized in Table B.1 and Table B.2 in Appendix B.2. Specifically, Column (1) of Table B.1 presents the estimation results when the photo quality is assessed not through the labeled score but objective features such as the contrast and the vibrance of the picture. Column (2) further includes a second-order term for the objective photo score to capture the potential nonlinear effect of the photo quality. We repeat the same estimation in Columns (3) and (4) using the labeled photo quality. All the specifications are estimated through GLM. The details for all the setups are presented in Appendix B.2.

Furthermore, we examine how photo type misspecification and misspecified listing status could bias our estimation results. While our model achieves an overall accuracy of 84%, Figure 3-2 indicates that certain room types, such as bedroom, are prone to misclassification. We thus use synthetic data to investigate how photo type misclassification affects our estimation results. Meanwhile, the consideration set may not be strictly shrinking over time, a new apartment may pop up at any time, or an existing listing may switch its status between "B" (blocked) and "A" (available). To study the impact of misspecified listing status on the estimation results, we first examine the percentage of miscounted incidences through a separate dataset that contains the three-month trajectory of the status for each listing. We identify that such miscounted incidences account only for 1.24% of the total number of cases. We then investigate the impact of miscounted cases on our estimation results through synthetic data. The elaborate setups for the two abovementioned cases are presented in Appendix B.2, and the estimation results are shown in Table B.2.

3.4 Optimization and Counterfactual Analysis

In this section, we explore how our estimation results from Section 3.3.1 lead to implications for Airbnb's listing owners. Specifically, we are interested in the benefits associated with adjusting the layout of the Airbnb listing photo. To this end, we

formulate the layout optimization problem to obtain the optimal photo layout that maximizes the attractiveness of a listing. We then investigate the impact of the optimal photo layout on the revenue for each listing.

To pave the way for our counterfactual analysis, we first determine the optimal photo layout for each listing. From Table 3.7, we observe that posting multiple bedroom (living room) images leads to decreasing (increasing) marginal returns, and therefore, selecting the optimal layout is nontrivial (i.e., it may not be optimal for owners to rely on their highest quality images if they duplicate room types) and requires an optimization framework. To this end, we first reformulate the impact from the photo layout, $\sum_m \left(\alpha_m^c s_m^c + \alpha_m^{nc} S_m^{nc} + \delta_m S_m^{nc} \mathbb{1}_{\{K_m^c=1\}} \mathbb{1}_{\{K_m^{nc}\geq 1\}}\right)$, in matrix representation. Suppose a listing owner has a set of N images with scores $\{s_n\} \subseteq \{[1,7]\}^N$, we use H, a $N \times 5$ matrix, to summarize the quality score as well as the room type of each image. Each row in H corresponds to an image, and each column corresponds to a room type ($\{B, L, O, T, K\}$). An entry only takes a positive value of its image quality score if the image is of the room type designated for the column. In other words, $H_{nm} = s_n \mathbb{1}_{\{Type_n=m\}}$. From the data input H, the listing owner must decide which image to use as the cover and which images to use in the four non-cover spots. Denote Z to be a $N \times 5$ decision variable matrix, where $z_{ni} = 1$ if the nth photo is selected to for the ith slot ($i = 1$ means covers). Also, through the previous estimation procedures, we obtain the coefficients on the layout impacts α_m^c, α_m^{nc} and δ_m. The integer programming formulation in matrix representation is as follows:

$$
\begin{aligned}
\max_{\boldsymbol{Z}} \quad & Z_1^T H \alpha^c + \sum_{i=2}^{5} Z_i^T H \alpha^{nc} + \sum_{i=2}^{5} Z_i^T H \delta \; \mathbb{1}_{\{Z_1^T H (Z_j^T H)^T > 0\}} && (3.15)\\
\text{s.t.} \quad & \sum_{n=1}^{N} z_{ni} = 1 && \forall i = 1, ..., 5 \\
& \sum_{i=1}^{5} z_{ni} \leq 1 && \forall n = 1, ..., N \\
& z_{ni} \in \{0, 1\} && \forall n, i,
\end{aligned}
$$

where α^c, α^{nc} and δ are the vectors of estimated parameters, and Z_j is the jth

column of the decision variable matrix Z, corresponding to the index of images. The objective function follows the definition of photo layout impact specified in Equation (3.1): $Z_1^T H\alpha^c$ and $Z_j^T H\alpha^{nc}$, $j = 2, ..., 5$, capture the impact of the cover image and the non-cover images, respectively. The indicator function $\mathbb{1}_{\{Z_1^T H(Z_j^T H)^T > 0\}}$ specifies whether a non-cover image is of the same room type as the cover; thus, the third term captures the potential impact of duplicate images. The first constraint in Equation (3.15) implies that each slot can only display one image, and the second constraint is that each image can be only displayed in at most one slot. This formulation is nonlinear since it involves several indicator functions $\mathbb{1}_{\{Z_1^T H(Z_j^T H)^T > 0\}}$, determining if the image at slot j has the same type as the cover's (the first slot). Directly solving this nonlinear integer programming problem is challenging due to the large feasible region resulting from hosts having a large number of photos. We therefore propose an algorithm to solve for the optimal solution.

Algorithm 2:

input : Scores s_n, Types $Type_n$, Model parameters α^c, α^{nc} and δ

for $m \in \{B, L, O, T, K\}$ **do** `// Denote as scenario` m

  find $n^* = \arg\max_n (s_n | Type_n = m)$, compute $v_{n*} = s_{n^*}\alpha_m^c$;

  for $n \in \{1, ..., N\} \setminus \{n^*\}$, **do**

    if $Type_n = m$ **then** compute $v_n = s_n(\alpha_m^{nc} + \delta_m)$; `// Duplicate effect`

    else compute $v_n = s_n\alpha_m^{nc}$;

  end for

  For the sequence $\{v_n\}, n \in \{1, ..., N\} \setminus \{n^*\}$, find $n^i = \arg_n v_{(N-i+1)}, i = 1, 2, 3, 4$;

  Compute $V_m = v_{n^*} + \sum_{i=1}^4 v_{n^i}$

end for

find $m^* = \arg\max_m V_m$;

apply the optimal layout as in scenario m^*

Intuitively, this algorithm compares different scenarios, under each of which we place a room type image with the highest score into the cover slot. Once the room type of the cover image is fixed, we specify the impact of the four non-cover slots. Finally, we compare the total utility obtained from each scenario and select the one with the highest value. We state the optimality of Algorithm 2 as follows:

Proposition 3.3. *When* $\alpha_m^c \geq \alpha_m^{nc}$ *and* $\alpha_m^c \geq \alpha_m^{nc} + \delta_m$, $\forall m \in \{B, L, O, T, K\}$,

Algorithm 1 recovers the optimal photo layout.

Proposition 3.3 holds when the cover image has a larger impact on customer utility than the non-cover ones, with or without the duplicate effect, which our estimation results in Table 3.7 confirm. We present the proof of Proposition 3.3 in Appendix B.4.

Applying the estimation results in Table 3.7 to Algorithm 2, we obtain the optimal photo layouts for listings across the ten neighborhoods. We contrast the differences in the room type and photo quality of cover images between the current and the optimal photo layouts in Figure 3-5. Panel (a) of Figure 3-5 shows that, in optimality, more than 71% of the listings should use a bedroom photo as the cover image. Our estimation results suggest high-quality bedroom cover images lead to the most significant improvements in a listing's demand. On the other hand, very few listings use photos of the kitchen and outside views as cover images since such photos usually do not contain critical information about the listings. Panel (b) of Figure 3-5 compares the quality of cover images in the current photo assortment and the optimal one. The qualities of the cover images increase for all room types under the optimal photo assortment, indicating that when switching to the optimal assortment, listings do not only benefit from a higher impact from the cover photos with the right room types but also an increase in photo quality.

We then compare the statistics for all 2,561 listings (2,047 non-superhosts and 514 superhosts) over all of Manhattan. Before the optimization, the means of the estimated layout effects are 0.816 and 0.846 for non-superhosts and superhosts, respectively. The two-sample t-test has a p-value of 8.6×10^{-4}, suggesting that superhosts adopt better photo layouts on average. After optimizing the layouts, the same result holds: the average layout impacts for non-superhosts and superhosts are 0.947 and 0.989, respectively. In summary, superhosts tend to post photos with a higher overall impact, potentially because (1) superhosts tend to put more effort into selecting photos to post, and (2) because superhosts, on average, post higher numbers of photos than non-superhosts (18.5 vs. 15.2), they have more photos to choose from when maximizing the impact of first five images. We also test whether superhosts are

more likely to post the optimal photo suggested by our counterfactual analysis as the cover photo. We find that while 15.7% of the non-superhosts had cover photos that conformed to the optimal solution suggested by our counterfactual analysis, 18.3% of the superhosts had already adopted the optimal cover photo.

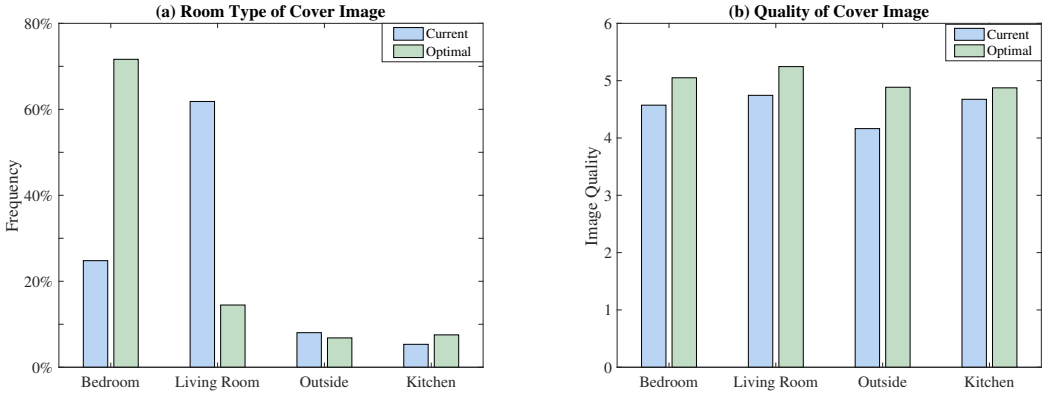

Figure 3-5: Quality and Room Type of Cover Image under the Optimal Photo Layout.

It is worth pointing out that listing owners on Airbnb enjoy a high degree of autonomy in that they can freely decide the quality and sequence of the images on the listing's webpage. However, it is still of interest to Airbnb to refine its current image-posting guidelines [10] and provide recommendations on the layout of the images to listing owners because those who follow the guidelines will enjoy higher demand and thus generate higher revenue for the platform.

After we obtain the optimal photo layouts for all these listings, we conduct three counterfactual analyses to study how the optimal photo layouts affect listings' rental probabilities and revenue.

Impact of Optimal Photo Layout on Rental Probability (Individual Level). We first study the impact on rental probability when a listing unilaterally switches to the optimal photo layout suggested by Algorithm 2. Under optimality, the impact from photos for listing i improves from the current value, V_i, to V_i^*, which is

$$V_i^* = \sum_m \left(\hat{\alpha}_m^c s_{mi}^{*c} + \hat{\alpha}_m^{nc} S_{mi}^{*nc} + \hat{\delta}_m S_{mi}^{*nc} \mathbb{1}_{\{K_{mi}^c = 1\}} \mathbb{1}_{\{K_{mi}^{nc} \geq 1\}} \right) , \tag{3.16}$$

[10] https://www.airbnb.com/help/article/746/how-can-i-take-great-photos-of-my-listing

where the values of $\hat{\alpha}_m^c$, $\hat{\alpha}_m^{nc}$ and $\hat{\delta}_m$ are presented in Table 3.7, and s_{mi}^{*c} and S_{mi}^{*nc} are the quality scores of photos in the optimal layout. In addition, the utility obtained by customer k from booking property i increases from the current value μ_{ik} to $\mu_{ik}^* = \beta_p p_i + V_i^* + \boldsymbol{\tau} PROPERTY_i + \boldsymbol{\gamma} CONTROL_i + \epsilon_{ik}$, with the rest of the listing features remain the same.

After we derive the mean utility for each property, we calculate the average increase in booking probability across listings as follows:

$$\Delta\mathbb{P} = \frac{1}{ND} \sum_{i=1}^{N} \sum_{d=1}^{D} \frac{\mathbb{P}_d(\mu_i^*, \mu_{-i}, B_d) - \mathbb{P}_d(\mu_i, \mu_{-i}, B_d)}{\mathbb{P}_d(\mu_i, \mu_{-i}, B_d)}. \tag{3.17}$$

In Equation (3.17), for a given day d, we assume that the demand (i.e. the total number of listings booked, B_d) is identical to what we observe in the data. Additionally, we denote $\mathbb{P}_d(\mu_i, \mu_{-i}, B_d)$ as the simulated booking probability for listing i with utility μ_i on day d, given the total number of bookings, B_d, and the utilities for other listings, μ_{-i}. Similarly, $\mathbb{P}_d(\mu_i^*, \mu_{-i}, B_d)$ is the probability that listing i gets booked on day d under the optimal photo layout. We calculate the renting probability of listing i, $\mathbb{P}_d(\mu_i, \mu_{-i}, B_d)$ using a 10,000 iteration sampling-without-replacement simulation, in which the sampling probability for each listing is proportional to $\mathbb{E}[e^{\mu_{ik}}]$ as $\epsilon_{ik} \sim Gumbel(0, 1)$. Through a similar simulation, $\mathbb{P}_d(\mu_i^*, \mu_{-i}, B_d)$ is obtained when listing i unilaterally adopts the optimal photo layout, in which case the increased attractiveness for listing i also lowers the relative booking probabilities for the rest of the listings. Repeating the analysis to all the listings in our group, we find that optimal photo layout, on average, increases the booking probability by 11.0%.

Notably, the real average impact when listing i unilaterally adopts the optimal photo layout can be higher than $\Delta\mathbb{P}$ for two reasons. The first is the spillover effect. The optimal photo layout first leads to increases in booking probability. Then, the increased bookings translate into a higher number of reviews over time, which, according to our estimation results in Table 3.7, further boosts the booking probability, and the positive feedback cycle continues. Hence, the impact of the photo layout will increase over time and will impact demand through multiple channels.

The second reason is that, according to Table B.2, the misclassification of the room type can result in our estimation providing an attenuated magnitude for the impact of the cover image. As a result, failing to account for such misclassification leads our counterfactual analysis to understate revenue gain when each listing switches to the optimal photo layout, as the increase in the overall attractiveness from the optimal photo layout should be higher under the correctly specified scenario.

Monetary Value of Optimal Photo Layout (Individual Level). We are also interested in the gain from the optimal photo layout to the equivalent dollar value. To this end, we study the following questions: Assuming listing owners display the optimal collection of photos suggested by Algorithm 2, how much can the owners raise their rental prices to neutralize the gain from displaying the optimal photo layout? What is the corresponding gain in revenue from the increase in rental rates? When we switch from the current photo layout to the optimal one, the photo impact improves from V_k to V_k^*. Since the attractiveness of listing i perceived by customer k can be written as $\mu_{ik} = \beta_p p_i + V_i + \boldsymbol{\tau} PROPERTY_i + \boldsymbol{\gamma} CONTROL_i + \epsilon_{ik}$, to keep the μ_{ik} the same, the new price will increase to:

$$p_i^* = p_i + \frac{V_i - V_i^*}{\beta}\,. \tag{3.18}$$

Following Equation (3.18), the counterfactual analysis suggests that by displaying the optimal photo layout, listing owners can, on average, improve the annual revenue by \$1247.6 while maintaining the same demand. The magnitude of revenue gain is substantial, especially considering that the improvement is achieved by adjusting the sequence of existing photos on the website, which incurs no cost at all.

Impact of Optimal Photo Layout on Listing Demand and Revenue (Platform Level). When every listing changes to the optimal photo layout simultaneously, the calculation of the revenue gain for each listing is complicated by the fact that (1) improvements across all Airbnb listings would intensify the competition among them and could result in decreased demand for certain listings; and (2) Airbnb would become more attractive as a platform after the adoption and would bring in new cus-

Algorithm 3:

input : Arrival list $\mathcal{M}$, Constant list $\mathcal{C}$, True parameters β_0, Features X_i, True Demand b_0

$b_{sim} = [0, 0, ..., 0]$ // Initialize the number of simulated bookings
$S^a = [1, 2, ..., I]$ // Initialize the availability set
for $M \in \mathcal{M}$ **do** // Each value of possible number of arrivals
  for $c \in \mathcal{C}$ **do** // Each value of constant c
    for $k \in \{1, ..., M\}$ **do** // Each customer arrival
      Realize utility for each option: $u_i = \beta_0 X_i + c + \epsilon_i$, $u_0 = \epsilon_0$
      Pick $i^* = \arg\max_{i \in \{0,1,...I\}} u_i$
      if $i^* \neq 0$ **then** // Not the outside option
        $b_{sim}[i^*] = b_{sim}[i^*] + 1$
        $S^a = S^a \setminus i^*$ // Listing i^* become unavailable
      end if
    end for
    Compute $err_{c,M} = \sum_i (b_{sim} - b_0)^2$
  end for
  Pick $c^*_M = \arg\min_{c \in \mathcal{C}} err_{c,M}$
end for

output: Simulated bookings under c^*_M for each M

tomers who previously chose to stay at hotels or alternative lodging establishments.

Thus, to study the impact of optimal photo layout adoption at the platform level, i.e., when every host simultaneously switches to the optimal photo layout, we need to account for the impacts of the intensified competition within the Airbnb platform and the increased market share of Airbnb. We assume that market size (total number of customers that have lodging needs) is known. We consider four scenarios in which the market size is 10, 5, 1, or 0.5 times as large as the number of listings. For each market size M, we estimate c, the constant utility value for the outside option, by minimizing the discrepancy between the simulated number of bookings and the actual number of bookings documented in the data. As we have shown through our synthetic data that the PCM estimation results are robust in the absence of outside option data, we use the results in Table 3.7 as the underlying parameters for customer utility in our estimation process. After deriving the constant c, we update the utility for each listing i by changing the impact of the photo layout from V_i to V_i^* and simulate the number of bookings in the presence of the outside option to calculate the demand and revenue gains. Algorithm 3 summarizes our procedures for our counterfactual

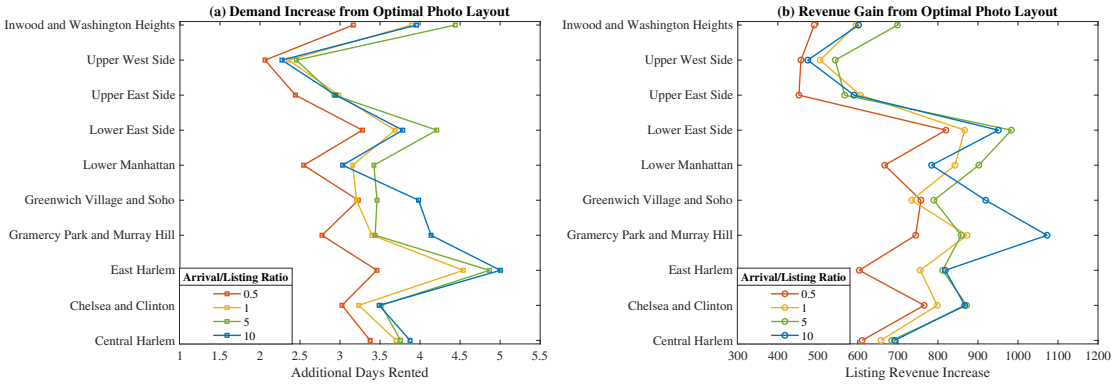

Figure 3-6: Counterfactual Analysis. (a) The demand increase from optimal layout; (b) The revenue gain from optimal photo layout.

analysis when all the listings simultaneously change to the optimal photo layout.

Figure 3-6 presents the by-region improvements when all listings on Airbnb simultaneously switch to the optimal photo layout. Depending on the specific neighborhood and market size, listings, on average, will be booked for two to five more days in a year, which boosts the revenue by \$500 to \$1100, respectively. As the constant utility, c of the outside option, is different under different market sizes, we note that a larger market size may not induce a higher number of transactions or more significant revenue gains. Additionally, when all the listings update their photo layouts, the revenue gain is, on average, smaller than the case with a single unilateral change to the optimal photo layout, suggesting that the intensified competition outweighs the benefit additional customers bring in through improved platform attractiveness.

Chapter 4

Large-scale Price Optimization for an Online Fashion Retailer

4.1 Introduction

Markdown and promotional pricing have been popular in the operations management field, where researchers and companies try to understand consumers' purchasing behavior by building demand forecasts and optimizing markdowns. The online fashion retailing environment has some unique features and challenges within the field. Firstly, it is common for a company to manage a large number (typically hundreds of thousands) of products across different markets and product categories in the online retailing environment. The scale of the problem imposes requirements on the efficiency of the optimization framework as it must be executed on a weekly or even daily basis. To make things more complicated, business targets exist that tie individual products together. For instance, from the business planning perspective, the company may set a particular revenue target for specific product categories or individual countries, which is a constraint that applies to all the products within the category and/or country. The challenge is that jointly optimizing all the products may be intractable, and decomposition methods must be applied. Another challenge is that we would have specific business constraints that require careful modeling, even for single-product discount optimization. For example, due to the fear of confusing

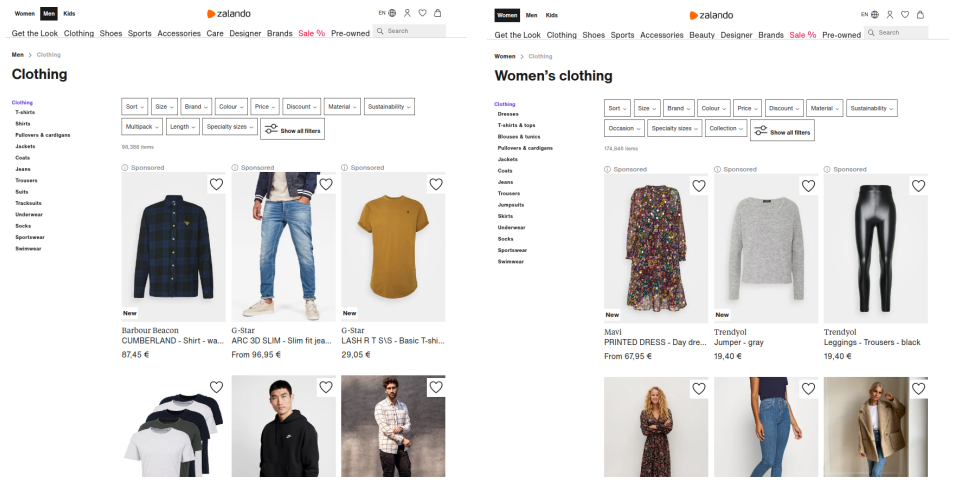

Figure 4-1: Zalando's homepages for the "Men" (left) and "Women" (right) segments.

customers or receiving negative customer responses, discounts on a single product should not change drastically from week to week. If these business constraints lead to a non-linear optimization model even for single products, the corresponding "global" problem with multiple products will have a higher level of complexity and would be impossible to solve. Finally, due to the nature of the fashion industry and the large scale of products, there is a long tail of products with very few or no historical sales data. This makes both demand forecast and optimization challenging.

In this work, we collaborate with Zalando (`www.zalando.de`), Europe's leading online fashion platform, which delivers products to more than 28 million active customers in 17 countries. On the fashion platform, customers can find a wide assortment of around 600,000 articles from more than $2,500$ brands for 6.5 billion euros yearly sales as of 2019. Zalando offers customers a comprehensive selection of apparel, footwear, cosmetics, and accessories for women, men, and children with free shipping and return shipping. A screenshot of the web page is presented in Figure 4-1. Zalando's logistics network with five central logistics centers enables efficient delivery to all customers throughout Europe, supported by local distribution focused on local customer needs in northern Italy, France, and Sweden. The fashion, technology, and logistics triad offers added value to the customers and brand partners.

The goal of the collaboration is to manage the prices of products through a sales season so that profit is maximized and overstock at the end of the season is minimized.

Additionally, the process aims to distribute discounts so that business targets (e.g., growth in a particular market) are reached. The implementation of this process in the past has been a hybrid of an automated optimization process, which recommends profit optimal discounts for individual products, and a manual selection process from these discount recommendations. The manual selection process builds on the experience of pricing experts for the individual markets, which use their expertise to select the discounts so that the market's business targets are reached.

There are a few challenges in the manual process. First, as Zalando's business grows, scaling this process is not easy. Second, because Zalando has multiple targets by product category and country level, intuition and experience are insufficient. Finally, it is not clear that the current process generates effective discounts. Therefore, our objective is to develop a fully automated system that satisfies the various business requirements and allows Zalando to maximize its impact on the bottom line.

Like most online retailers, Zalando has a regular price update cycle to ensure that each product (or Stock Keeping Units, SKU) receives an optimal discount. Weekly price updates are typically synchronized between all markets, while prices in different countries (or markets) are set independently. The shop operates with a global stock assumption that every product can be sold in any country without limitations by sharing a shared inventory pool. This means that we do not have direct control over the sales: it is not an option to turn down a customer to reserve the product for customers in other countries. We can only impact sales through appropriate discounts.

Weekly selection of prices for such a large assortment is near impossible to do manually. That is why the first generation of the automatic price recommendation system was introduced in Zalando, aiming to maximize total profit. The insights were to discount heavily the SKUs that are expected to have large overstock (e.g., due to optimistic buying decisions or drastic changes in season's weather) and to discount conservatively with SKUs that are selling well in the shop. Such an automatic system proved to be beneficial and led to significant increases in financial indicators, and contributed to the company's successful growth.

However, such a system lacks an important feature, crucial for a large-scale multinational business. Having a drastic financial impact, pure discounting does not allow to steer towards specific financial targets. It does not allow easily to include strategic company goals in the pricing process. This leads to a situation where discounts produced by the system still need manual intervention from commercial planners to steer discounts toward financial targets. In this paper, we also refer to the latter as "global" constraints since our model will have to solve at a more "global level" instead of by single products.

The increase in manual efforts reveals a need for a new generation of pricing systems, which we call the "target steering" system. Such a system aims not only to make price recommendations for overstock mitigation but also to provide discounts that satisfy weekly financial targets (like certain revenue levels in a selection of countries). This paper describes the design and implementation details of such a large-scale price recommendation system, which proved to be a challenging task on such a large scale.

Given the demand forecasts under a collection of discrete prices, the objective of Zalando is to find the discount levels that maximize the total profit over the entire selling horizon while taking into account both "local" and "global" business constraints. On the "local" single SKU level, due to operational sharing inventory pool assumption, Zalando needs to balance the sales across different countries. Specifically, we have to make sure that no inventories are reserved strategically for demands in other countries or later time periods (this is referred to as "stock hedging"). To address this challenge, we formulate the price optimization problem as a mixed-integer programming (MIP) problem and incorporate the balancing constraints to solve the "stock hedging" issue. In addition, since our forecaster is at the size-aggregated SKU level, due to each SKU's limited availability of different sizes under low stock, we propose a new stock-dependent method to adjust the demand forecasts and provide more accurate input into the pricing system.

On the "global" level, when we jointly optimize the discounts of different SKUs in selected categories, the total revenue and weighted average discounts of the SKUs should meet specific "global" steering targets. This is because pricing managers

need to manage the discounts across articles to deliver the forecasted discount spend and achieve the revenue targets. The weighted average discounts (or, say “discount spend") are essential for pricing managers' planning decisions. To reduce the computational complexity raised by these global constraints, we decompose the global problem into parallel sub-problems of each single SKU by using Lagrangian decomposition and propose an efficient algorithm to find the optimal Lagrangian multipliers.

Finally, we developed an aggregation framework to cluster SKUs based on categories and similarities to reduce computational complexity. This framework enables solving the problem via a 3-step process. In the first step, we solve the aggregated problem in a way that decouples the problem into a problem for each product category. In the second step, we apply the Lagrangian method to decouple the problem for each SKU. Finally, in the last step, we solve a large number of SKU-specific problems. Notably, the aggregation method enables planners to test various pricing strategies before finalizing the season's targets for each product category and country.

We also conduct field experiments to validate the optimization system in real business cases. In both the online and offline environment, we design and implement experiments to validate whether the proposed framework is capable of steering towards specific business targets. The results in the target group satisfied the targets, where global constraints are met, and the weighted average discount values are close or within the target bounds most of the time. Previously, the commercial team managed different global constraints manually using intuitions and heuristic processes. This new methodology will automate the process of pricing, reducing the manual work and making scaling up much more manageable.

4.2 Literature Review

Promotional pricing is a sales strategy in which brands temporarily reduce the price of a product or service to attract prospects and customers. By lowering the price for a short time, a brand artificially increases the value of a product or service by creating a sense of scarcity. Consumer-goods companies and retailers realize that

getting pricing, markdowns, and promotions right across all brands and channels is critical to survive and thrive. In-depth overviews of this literature can be found in Talluri and Van Ryzin (2006) and Özer et al. (2012). Previous research contributes to developing and implementing pricing decision support tools for retailers. Smith and Achabal (1998) developed clearance prices and inventory management policies. Natter et al. (2007) implemented a decision-support system for dynamic retail pricing and promotion planning. In the meantime, more and more companies have adopted industry software to facilitate their pricing decisions. For instance, software companies like LOKAD and BlueYonder provide software systems solutions to demand prediction and optimization within the supply chains. However, such products may not be able to tailor to the company's needs fully, and in many cases company demands its system of forecast and optimization. Making the right pricing decision is a critical step for modern companies to succeed and enable smart decision-making, taking advantage of the visibility of data (N. Boute et al. 2020).

Our work is to apply revenue management techniques for optimizing prices in the online fashion retailing environment across different countries under inventory constraints. The problem we study in this paper includes the following distinguishing features that define our unique position in the revenue management literature.

First, in the online retailing environment, the richness and availability of data enable us to build demand forecasts using historical observations. This is related to the online learning setting for price optimization. There is an increasing stream of research on online learning models, which often assume an unknown linear demand model. den Boer (2015) provides a comprehensive survey on this topic. Papers in this research stream study the fundamental trade-off between experimenting to improve estimates of the unknown demand model (the exploration) and leveraging current estimates to maximize revenue (the exploitation) (Bu et al. 2020). There are many papers that developed online pricing models under the assumption of linear demand model (e.g., Keskin and Zeevi 2014, den Boer 2015, Qiang and Bayati 2016, den Boer and Keskin 2019). Nambiar et al. (2019) and Ban and Keskin (2020) assumed generalized linear demand model. In this paper, Zalando builds a deterministic demand

function based on neural networks, making it powerful to incorporate all the internal and external data to improve prediction. However, it is highly nonlinear. This non-linearity makes the optimization model more challenging to deal with.

Second, the pricing process usually contains two key components: demand forecasting and price optimization. An essential input for the success of the price optimization model is the predicted demand values. Recent advances in machine learning techniques and richness of data have motivated innovative data-driven approaches to forecast demand and optimize price. For example, Ferreira et al. (2016) studied a pricing problem for an online flash fashion retailer, Rue la la. That paper applies random forests to estimate customer demand under different price levels and proposes an efficient optimization algorithm based on mixed-integer programs to make discount decisions. Caro and Gallien (2012) study a clearance pricing problem for fast-fashion retailer Zara. They build a demand forecast model to address the lack of price-sensitivity data in that work. They then feed the demand forecast to an optimization model to determine price markdowns. Cheung et al. (2017) study a promotion pricing problem for Groupon, a large e-commerce marketplace for daily deals. In that paper, they develop a pricing policy that dynamically learns customer demand using real-time sales data under limited price experimentation. Ma et al. (2018) applies random forest models to predict demands for a CPG company, followed by a pricing optimization model.

Third, another interesting feature of the problem is the large scale due to the nature of fashion products, often involving large assortments. Individual product is sold across multiple countries with a shared inventory pool, and the inventory allocation across different countries and different time periods need to be *balanced*. Namely, the company cannot reserve inventory for a specific country or time period. Furthermore, the pricing decisions of different products are tied together under specific business steering targets across countries and product categories. Therefore, instead of solving the problem by articles, the model should be able to solve multiple products at a much large scale across countries within a reasonable time. There are very few papers resolving this challenge in the literature.

Finally, this work is also related to the literature investigating the operational challenges in the online fashion retailing industry. Apart from the several papers mentioned above, there has been more and more empirical work in the context of fashion retailing. Caro and Martínez-de Albéniz (2015) provides a comprehensive overview of the business models for the fast-fashion industry. Boada-Collado and Martínez-de Albéniz (2020) examines the impact of inventory levels on demand in the fashion retailing setting. Fisher et al. (2018) validate the pricing competition model in the online retailing setup through field experiments. Our paper combines data-driven approaches and optimization modeling methods, validated through offline tests with historical data and real-world field experiments.

4.3 Demand Forecast Model

Before introducing the price optimization framework, we first give a brief overview of the demand forecast model, generating deterministic demand predictions under all price levels and serving as the optimization model's input. To make the price optimization model succeed, the accuracy of the demand forecaster is critical. There are many commercial promotion software available in the market. However, they either can not meet the high requirement of accuracy or are limited to certain types of businesses environment. Zalando has its unique challenges (large scale assortment, long tail of low sales products, operating in multiple countries, long planning horizon, etc.), making it essential to develop our own operational processes/models.

Zalando uses a collection of forecast models that provide (size aggregated) article-level forecasts every week. The forecaster is based on the Transformer architecture Vaswani et al. (2017) with some adjustments to make it suitable for time series forecasting. The transformer is a recently developed machine learning tool based on neural networks. It has been widely used in natural language processing and achieves outstanding performance. In Zalando's context, Transformer inputs the sales history of every product in all countries and some product-specific feature information (e.g., brand, color, style, product category) to predict the demand for future weeks.

The model is retrained every week with new incoming data to predict, considering discount levels, past sales history, countries, weeks, and product-specific features mentioned above. It uses weighted least square error as the measure of accuracy. It has been tested to outperform the previous forecast models Zalando was using, including gradient-based auto-regressive models, random forests, and other types of neural networks. We present the relative performance metrics in Table 4.1, where ASF4 is the name for the gradient-based auto-regressive benchmark model. LSTM is another type of neural network model. We mask the percentage error values so that ASF4 has a calibrated level of zero, then present the absolute percentage difference in terms of errors for the other two models. The table includes both first-week forecast accuracy and aggregate level forecast accuracy for the whole selling season. It is observed that the Transformer model obtains not only the lowest error with massive improvement from the other two models but also the most minor bias, which is essential for the optimization procedure.

Table 4.1: Performance Comparisons between Demand Forecast Models

Models	1st Week Error	Seasonal Error	Bias
ASF4	0 (calibrated)	0 (calibrated)	14%
LSTM	-30%	+9%	12%
Transformer	-30%	-9%	-3%

In business practice, we have the following two observations on demand. First, since markets are based in different countries, it is unlikely that demand in one country will depend on the prices in other countries. It is observed that same applies to different time periods and different products. The assumption is that the forecast in one country does not depend on discounts from other countries, nor should the forecast for one week rely on the forecast in other weeks. This assumption allows us to characterize demand independently across countries and time periods in the Transformer architecture. The second observation is that discount levels are not continuous, and usually take discount steps of 5% (e.g., 15%, 50%). This allows us to model the optimization problem into an MIP to select the optimal discount among the discrete discount levels. For this MIP, we require the forecaster to tabulate the

predicted demand under every possible discount level for every country and every coming week into large tables, which will feed as inputs to the optimization model in the following section.

4.4 Single SKU Discount Optimization

We start with the price optimization model for each single SKU. Recall that each SKU's discount needs to be optimized for 17 countries (denoted as C) and generally a season of 40 weeks (denoted as T). The forecaster in the previous section will generate the predicted demand for each discount level on the price ladder (denoted by L). Given L discrete discount levels, we need to decide the optimal discount for each product over a selling horizon of T weeks across C countries. Throughout the paper, we use $[n]$ to denote the set $\{1, 2, \ldots, n\}$. Let P_c denote the original (undiscounted) price of the SKU in country $c \in [C]$, and let d_l be the discount value for discount level $l \in [L]$. Specifically, we have a total of $L = 15$ discount levels, ranging from 0% to 70% off, with a step size of 5%, in which $d_1 = 0$ denotes the undiscounted price and $d_L = 0.7$ denotes 70% off the original price.

Let $D_{c,t,l}$ and $R_{c,t,l}$ be the demand and return rate forecast, respectively, for country $c \in [C]$ in week $t \in [T]$ under discount level $l \in [L]$. We assume that demand $D_{c,t,l}$ and return rates $R_{c,t,l}$ are deterministic and provided as inputs to the optimization model. Given the sales in week t, the corresponding returns are distributed in the following weeks according to the return base vector RB. In other words, RB_1 fraction of the sales will be returned in the upcoming week. The RB vector is also assumed to be fixed and provided by Zalando as an input.

Our decisions are binary variables $z_{c,t,l} \in \{0, 1\}$, which indicate the choices of discount level l in country c on week t, and sales variables $x_{c,t,l}$, which characterizes the sales under discount level l in country c on week t. Let y_t be the stock level at the beginning of week t, and y_{end} be the stock leftover after selling horizon T. At the beginning of week t, we have stock replenishment A_t that is predetermined before the whole selling horizon. At the end of selling horizon, the remaining stock y_{end} has a

salvage value of SV per unit. Let $\pi_{c,t,l}$ be the profit of selling one item of the SKU in country c and week t under discount level l. [1]

To maximize the total profit of selling the product, we formulate the single SKU discount optimization problem into a MIP.

$$(P) \quad \max_{x,z} \quad \sum_{c,t,l} \pi_{c,t,l}\, x_{c,t,l} + y_{end}\, SV \tag{4.2}$$

$$s.t. \quad y_t = y_1 - \sum_{c,s<t,l} x_{c,s,l} + \sum_t A_t + \sum_{c,s<t,l} (\sum_{i=1}^{t-s} RB_i) R_{c,s,l} x_{c,s,l} \quad \forall t = 2, ..., T \tag{4.3}$$

$$\sum_{c,l} x_{c,t,l} \leq y_t \qquad \forall t \tag{4.4}$$

$$x_{c,t,l} \leq z_{c,t,l} D_{c,t,l} \qquad \forall c, t, l \tag{4.5}$$

$$\sum_l z_{c,t,l} = 1, \qquad \forall c, t \tag{4.6}$$

$$z_{c,t,l} \in \{0, 1\}, x_{c,t,l} \geq 0 \qquad \forall c, t, l \tag{4.7}$$

The objective is to maximize the total profit, both in and after the selling season. Constraint (4.3) specifies the stock dynamics for each time period, where the stock level at the beginning of week t is equal to the initial stock y_1 minus sales, plus replenishment and returns from previous weeks. Constraint (4.4) requires that the total sales for a specific time period have to be less than or equal to the remaining

[1] Specifically, we have

$$\pi_{c,t,l} = \frac{1}{(1+\text{CCR})^{t/52}} \left(\frac{\sum_l P_c(1-d_l)(1-\text{CO})(1-R_{c,t,l})}{1+\text{VAT}_c} \right) - \frac{1}{(1+\text{CCR})^{t/52}} \left(R_{c,t,l}\text{CR}_c - \text{CF}_c \right) \tag{4.1}$$

where CCR and VAT are constants. CO, CR and CF are the coupon loss, return and fulfillment cost, respectively.

stock. Constraint (4.5) limits the sales variable for each c, t, l (country, time and discount level) combination. So that when $z_{c,t,l} = 0$, the sales must also be zero, and when $z_{c,t,l} = 1$, the sales will be less or equal to the forecast demand. Finally, Constraint (4.6) describes that only one discount level is allowed to be selected in each country and each week.

A natural question is that why we need to model the extra sales variable $x_{c,t,l}$, given that in reality we do not have control over it. We illustrate this point through a counterexample. Suppose that we do not model the sales variable, the simplified MIP will have the following form:

$$\begin{aligned}
\max_{z} \quad & \sum_{c,t,l} z_{c,t,l}\ p_l\ D_{c,t,l} & \\
s.t. \quad & \sum_{c,t,l} z_{c,t,l}\ D_{c,t,l} \leq Y & \\
& \sum_{l} z_{c,t,l} = 1 & \forall c, t \\
& z_{c,t,l} \in \{0, 1\} & \forall c, t, l
\end{aligned}$$

Suppose we have a toy model with a single period, single country, and two price levels $p_1 = 20, p_2 = 10$, so that $C = T = 1, L = 2$. The inventory level is $Y = 120$, and the demand forecaster gives us $D_{l=1} = 50$ and $D_{l=2} = 140$. In other words, if we set the high price of 20, we will have a demand of 50, and the low price of 10 will yield a demand of 140. With the simplified MIP, the optimal solution is to set the price high $p = p_1 = 20$, and the total profit is 1,000. However, in reality, we could set the low price and only satisfy partial demand, with $p = p_2 = 10$ and a total profit of 1,200. From this counterexample, we can see the limitation of the simple MIP in terms of the flexibility to capture different sales levels. We can think of this limitation as a result of having discrete price levels since if prices are continuous, we could always set the correct price for the demand just to deplete all the stock. As a result, we need to model sales $x_{c,t,l}$ as a decision variable in the model.

4.4.1 Business Constraints

The price optimization model (P) is a basic model that captures the stock dynamics and establishes the relationship between discount decisions and sales. From the business perspective, it is necessary to set certain limitations on the discounts in this basic model. The motivation could be either to avoid lousy customer perceptions and experience, adjust to specific promotional sales events, or simply from the business requirements. We summarize several types of single SKU business constraints as follows.

Minimum/Maximum Discounts. To allow flexibility of discount levels, the discount range for the basic formulation is set to be from 0% to 70%. However, in practice, the allowable discount range could be much narrower. Some limitations could come from important brand agreements associated with brand images; others can come from specific country (market) regulations. For example, we would not expect a newly on-shelf fashion product to have deep discounts. The model includes country-week specific minimum/maximum discount bounds to capture this constraint to restrict the discount ranges.

Maximum Upward/Downward Steps. Intuitively, customers will get upset if they find the price changes drastically within a short amount of time. In addition, significant discount increases may lead to an explosion of returns, which will incur non-trivial costs on the business. In other words, discount differences in consecutive weeks should not be substantial. We could upper bound these discount differences by country-week specific maximum upward/downward steps. For example, if one SKU has a discount of 20% in week one, and the maximum allowable upward and downward step are both 15%, then the feasible discount range for week two will be from 5% to 35%.

Discount Barriers. When discount levels are too marginal (like 5%), customers' perception of the promotion may be compromised. To address this issue, we impose country-week specific discount barriers so that the discount values are either 0% or above these specific discount barriers. For example, by setting a discount barrier of

20%, we disallow discount levels 5%, 10% and 15%.

4.4.2 Stock Hedging

In the previous section, we assume in model (P) that we can optimize over sales via decision variables $x_{c,t,l}$. Although this assumption allows a linear formulation of (P), it may cause "stock hedging" problems in the final solution, meaning the product's inventory is reserved for specific countries and certain weeks, which violates the operations in practice. The online shop operates with global stock, accessible by all countries, and it is not typical to reject sales from a specific country, even if it may be profitable. For example, a customer in Germany arrives in week one when stocks are available. The model (P) may reject her demand by setting $x = 0$ because it is more profitable to sell this inventory unit in Spain or week two. In reality, Zalando does not have the flexibility to turn down customers and "hedge" the stock, so the optimal solution from (P) is often not practical. We need to integrate further constraints to deal with the stock hedging problem.

Specifically, we go through the sales dynamics in each week as follows, and there will be two possible scenarios. When the inventory is sufficient in week t, i.e., $y_t \geq \sum_{c,l} z_{c,t,l} D_{c,t,l}$, it requires $x_{c,t,l} = z_{c,t,l} D_{c,t,l}$, i.e., sales equal to demand $x_{c,t,l} = D_{c,t,l}$ in each country for the selected price levels l with $z_{c,t,l} = 1$. The other scenario happens when the inventory is insufficient in week t, i.e., $y_t \leq \sum_{c,l} z_{c,t,l} D_{c,t,l}$, as a result not all the demand in week t will be satisfied. Here we assume that the customer arrival process in all countries are evenly distributed across the week, and we will have $x_{c,t,l} = y_t \cdot \left(z_{c,t,l} D_{c,t,l} / \sum_{c,l} z_{c,t,l} D_{c,t,l} \right)$, i.e., sales split proportionally to demand in each country for the selected price levels.

Combining the above two scenarios, the realized sales in practice can be described by the following equation:

$$x_{c,t,l} = z_{c,t,l} D_{c,t,l} \min\{1, \frac{y_t}{\sum_{c,l} z_{c,t,l} D_{c,t,l}}\}. \tag{4.8}$$

In our model formulation (P), we assume sales are also decision variables, and

simply put constraints $x_{c,t,l} \leq z_{c,t,l} D_{c,t,l}$ on the sales decisions $x_{c,t,l}$. This formulation allows an extra degree of freedom to allocate the stock across different week and countries. However, the optimal solution might reserve the inventory for specific countries and certain weeks with insufficient inventory, as the example above illustrates. If this is the case, we will see in the final solution that $x_{c,t,l} < D_{c,t,l}$ even when y_t can satisfy the demand of all the countries in the corresponding week. This violates the real-world sales pattern in (4.8), and we refer to this violation as the "stock hedging" problem.

We refer to the requirement in (4.8) as the sales-balancing conditions. Since restrictions are nonlinear, we cannot directly integrate them into our basic MIP formulation (P). To address this problem, We break down the conditions into country-balancing and week-balancing conditions and then add constraints into our formulation to capture these two conditions separately.

Country balancing constraints. The stock hedging problem across countries, the issue occurs when unbalanced sales occur across countries. For example, suppose we only have two markets, and the demand is 100 in Germany and 50 in Spain, we would expect the realized sales to be also 2:1. In other words, if inventory is sufficient (larger than 150), the sales will be equal to the demand. If the inventory is insufficient (say 100), the demand will be satisfied proportionally (66 in Germany and 33 in Spain).

We address the stock hedging issues across different countries by adding the following constraint into formulation (P):

$$\sum_{l \in [L]} \frac{x_{c,t,l}}{D_{c,t,l}} = \sum_{l \in [L]} \frac{x_{1,t,l}}{D_{1,t,l}} \quad \text{for} \quad c \in [C] \setminus \{1\},\ t \in [T]. \tag{4.9}$$

Intuitively, this constraint requires that sales through all countries should follow the same depleting rate (proportional to their demands). It makes sure that given a set of demands and inventory levels, there's only one way to distribute the stock across countries, which is to distribute proportionally to their demands.

Week balancing constraints. For the stock hedging problem across weeks,

the issue occurs when the model reserves stock for future weeks, when there are still unsatisfied demands in the current week.

We address the stock hedging issues across different weeks by adding a set of big-M constraints. We introduce binary variable ϕ_t for each time period. Specifically, $\phi_t = 1$ denotes sufficient inventory in week t and $\phi_t = 0$ denotes insufficient inventory in week t. We then add the following constraints:

$$\sum_{c,l}(D_{c,t,l}z_{c,t,l} - x_{c,t,l}) \leq M_{demand} \cdot (1 - \phi_t) \qquad \forall t = 1, ..., T \tag{4.10}$$

$$\sum_{c,l} x_{c,t,l} \geq y_t - M_{stock} \cdot \phi_t \qquad \forall t = 1, ..., T, \tag{4.11}$$

where M_{demand} and M_{stock} are large constants that upper-bound the total demand and remaining inventory in each week, respectively. When $\phi_t = 1$, constraint (4.10) in combination with (4.5) forces sales to be equal to demand and (4.11) is relaxed. When $\phi_t = 0$, (4.10) is relaxed, and (4.11) in combination with constraint (4.4) forces sales to be equal to the remaining inventory.

Corollary 1. *Set of constraints* (4.9)*,* (4.10) *and* (4.11) *is identical to desired sales pattern described by* (4.8) *given the original problem formulation.*

The proof can be found in the Appendix. By adding constraints (4.9), (4.10) and (4.11), we fix the stock hedging problem and capture the observed sales pattern in (4.8), while still maintaining the linear structure for the optimization model.

4.4.3 Limited Size Availability

In the context of the fashion industry, there is a clear separation between articles on the unit and aggregated level: throughout the paper, we have been using the notion of SKU for configuration level stock keeping units (e.g., sneakers of a specific brand, including all sizes). An SKU can be managed at a lower level (e.g., the white T-shirt of a particular brand-sized “M"). In principle, it is possible to set prices on size levels. The latter is not typically done due to the drastic increase of the problem scale.

In model (P), we assume that each SKU represents the same product with different sizes. Both the demand forecast and the inventory of each SKU are given on the aggregated level of all sizes. In reality, however, specific sizes of the SKU might be unavailable when the stock level is low, which might raise discrepancies between the observed sales and the demand forecast values. For instance, consider an SKU that contains a white T-shirt of the same style with sizes S, M, and L. When the stock level is much higher than the demand, it is likely that all sizes are available in the requested quantities. When the stock level is low and close to the demand, specific sizes may not be in stock sufficiently, and thus the corresponding demand cannot be fulfilled entirely. In this case, the given demand forecast values overestimate the actual demand, and we need to scale down these values based on the product's stock level.

One way to adjust the demand forecast is to use a "stock response" function, which is defined as a function that maps the product's stock level to a smaller value to approximate inventories in reality. Specifically, given stock level Y_t, the stock response function outputs a multiplier $\mathrm{sr}_t(Y_t)$ that is multiplied to demand forecasts $D_{c,t,l}$. In practice, Zalando has been implementing an exponential stock response function as follow:

$$\mathrm{sr}_t(Y_t) = 1 - \exp\left(-\alpha \cdot (Y_t/\mathcal{N})^{\beta}\right), \tag{4.12}$$

where $\mathcal{N}$ is the cardinality (number of different sizes) of the SKU, and α and β are parameters fine-tuned by fitting the historical data. A graphical illustration is plotted as the solid curve in Figure 4-2. The intuition is that when the stock level is very high, all sizes are expected to be available. Therefore, the stock response factor is close to one. On the other hand, when the stock level is low, it is more likely for specific sizes to become unavailable and the corresponding demand cannot be satisfied; therefore, realized demand will reduce by multiplying a factor of the stock response value.

Piecewise Linear Approximation

One challenge of the approach above is that the stock response function is nonlinear and will be multiplied by the demand and decision variables in the formulation, which will also become non-linear. To deal with this challenge, we adopt the approximation algorithm in Kontogiorgis (2000) to approximate the stock response function with a piecewise linear function that has K segments. The detailed algorithm is provided in the appendix. Figure 4-2 shows the approximation Piecewise Linear Interpolate (PLI) result with three segments, compared to the benchmark $P2$ two-piece approximation.

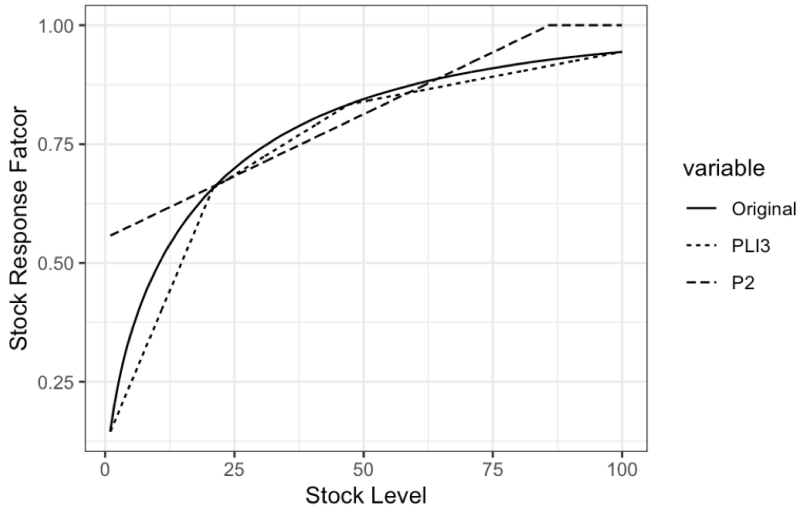

Figure 4-2: Stock response function and the piecewise linear approximation

There is a natural trade-off on how many pieces we should select for the approximation. We compare the maximum error, the running time of the model, and the objective difference for different approximation schemes. The result is presented in Table 4.2.

Table 4.2: Comparison between stock response approximation schemes

Methods	Max Error	Running Time (Seconds)	Objective Difference
P2	0.412	502	9.03%
PLI3	0.117	671	5.35%
PLI4	0.063	710	1.76%
PLI5	0.037	874	1.07%

Formulate Stock Response Factors

We explicitly formulate the stock response factors using the following constraints that describe the factor as a linear combination of the breakpoints of the piecewise linear curve. Let $0 = y^{(1)} \leq \ldots \leq y^{(K)} \leq y^{(K+1)} = \overline{y}_k$ be the breakpoints of the interval $[0, \overline{y}_k]$, and $f^{(1)}, \ldots, f^{(K)}, f^{(K+1)}$ the corresponding function value, i.e., $f^{(i)} = \text{sr}\left(y^{(i)}\right)$ for $i \in [K]$. Given inventory y_t at the beginning of week t, we have

$$y_t = \sum_{i=1}^{K+1} \mu_i y^{(i)} \tag{4.13}$$

$$\text{sr}_t = \sum_{i=1}^{K+1} \mu_i f^{(i)} \tag{4.14}$$

$$\sum_{i=1}^{K+1} \mu_i = 1 \tag{4.15}$$

$$0 \leq \mu_i \leq 1 \text{ for } i = 1, \ldots, K+1 \tag{4.16}$$

In addition, (4.13) and (4.14) should be a linear combination of **two consecutive breakpoints**. Let E_i with $i = 1, \ldots, K$ be the binary variables that indicate whether or not $y^{(i)}$ is selected as the left breakpoint. We then have constraints

$$\sum_{i=1}^{K} E_i = 1 \tag{4.17}$$

$$E_i \in \{0, 1\} \text{ for } i = 1, \ldots, K \tag{4.18}$$

$$\mu_1 \leq E_1 \tag{4.19}$$

$$\mu_i \leq E_{i-1} + E_i \text{ for } i = 2, \ldots, K \tag{4.20}$$

$$\mu_{K+1} \leq E_k. \tag{4.21}$$

4.4.4 Integrating Stock Hedging and Stock Response

We integrate extra constraints in the previous two subsections to deal with stock hedging and limited size availability issues, respectively. Unfortunately, the stock hedging behavior is impacted by the stock response constraint. It is assumed in stock

hedging that either the stock is depleted or the demand is fulfilled. However, with the stock response factor, the condition no longer holds. It should be updated that the stock is depleted or the demand modified by the stock response factor is fulfilled.

Due to the interference between stock hedging and limited size availability, we need to introduce an extra set of variables and constraints. Let $D'_{c,t,l}$ denote the demand forecast after the modification of the corresponding stock response factors, and we then have the following constraints that describes the modified demand forecast:

$$D'_{c,t,l} \leq \mathrm{sr}_t \cdot D_{c,t,l} \qquad \forall c,t,l \qquad (4.22)$$

$$D'_{c,t,l} \leq z_{c,t,l} \cdot D_{c,t,l} \qquad \forall c,t,l \qquad (4.23)$$

$$D'_{c,t,l} \geq \mathrm{sr}_t \cdot D_{c,t,l} - (1 - z_{c,t,l}) \cdot M_{\mathrm{sr}} \cdot D_{c,t,l} \qquad \forall c,t,l \qquad (4.24)$$

Given $D'_{c,t,l}$ the modified demand forecast, we need to update constraint (4.10) to

$$\sum_{c,l} D'_{c,t,l} - x_{c,t,l} \leq M_{demand} \cdot (1 - \phi_t), \quad \forall\, t \in [T] \qquad (4.25)$$

and additionally, constraint (4.5) to

$$x_{c,t,l} \leq D'_{c,t,l}, \quad \forall\, c,t,l. \qquad (4.26)$$

We now have the full-scale single SKU optimization problem that captures all relevant business constraints with the above constraints included. In summary, the constraints for the ingle SKU MIP are (4.3)-(4.7), (4.9), (4.13)-(4.21), and (4.22)-(4.26). The benefit of this linear formulation is the computational time. Solving each single SKU MIP with 17 countries, 40 weeks, and 15 price levels using a commercial solver will take several seconds. We can speed up the process by utilizing parallel computing power since, by far, each SKU is optimized individually. There are other approaches to speed up the single SKU optimization, and we will introduce an effective one in the next subsection.

4.4.5 Telescopic Method

We can aggregate on the time scale for single SKU optimization by using a telescopic method to reduce the computational time. The telescopic method is motivated by several observations in single SKU level optimization. First, the optimization model is solved only to obtain optimal discounts for the first week and will be resolved with new data inputs every week. As a result, the most relevant decisions will be the first-week discounts. Second, it is observed that demand and return forecasts for later weeks are less accurate, and intuitively we should "care less" about later weeks. Last, computational burdens are huge concerns in practice. In general, the optimization model needs to include the entire season that SKUs are offered to customers and could last for a whole year. In other words, the number of time periods can be as large as $T = 52$ weeks. Although one can resort to linear programming (LP) relaxations for faster computations, with the extra complexity of the formulation introduced in Section 4.4.2, 4.4.3, LP relaxations will be less reliable, and the computation time for solving MIP will explode as the number of weeks grows. In contrast, telescopic methods "combine" later weeks into fewer optimization periods, significantly speeding up the optimization routine.

Take an example where the optimization planning horizon is $T = 40$ if we introduce additional constraints $z_{c,t,l} = z_{c,20,l}$ for all $t \geq 20$, that is, forcing all the discounts after week 20 to be the same. The new optimal solution will be suboptimal for the original problem and yield a lower objective. Nevertheless, the optimal solution for the first week will not be far from the true optimal solution. Hence we can implement such an approximation by combining weeks after week 20 into one period.

Denote $\tau \in [t_1, ..., t_{\hat{T}}]$ to be the starting weeks for each period and we have $\hat{T}$ total telescopic periods, each with n_τ weeks. For example, $[1, 2, 7]$ means that we combine weeks 2-6 into one period and all weeks after (weeks 7-40) into another period. We denote the hatted version to be the telescopic updated version and update the model

inputs to their aggregated version:

$$\hat{D}_{c,\tau,l} = \sum_{i=t_\tau}^{t_{\tau+1}-1} D_{c,i,l} \qquad \hat{A}_\tau = \sum_{i=t_\tau}^{t_{\tau+1}-1} A_i \tag{4.27}$$

$$\hat{R}_{c,\tau,l} = \frac{1}{n_\tau} \sum_{i=t_\tau}^{t_{\tau+1}-1} R_{c,i,l} \qquad \hat{\pi}_{c,\tau,l} = \frac{1}{n_\tau} \sum_{i=t_\tau}^{t_{\tau+1}-1} \pi_{c,i,l} \tag{4.28}$$

Intuitively, we approximate demand, replenishment, and unit profit of a telescopic period by summing up the values of each week. For returns, we compute the average return rates of each week since these are fractional values. In addition, we approximate the business constraint inputs, including max/min discount levels, max upward/downward steps, and discount barriers, by taking the average across the corresponding weeks. The telescopic methods have proven to be an effective approximation heuristic and have been adopted by Zalando to speed up the weekly optimization routine.

4.5 Global Optimization

In previous sections, we formulated the discount optimization problem for single SKUs and addressed the potential issues caused by the linear MIP formulation. If each SKU behaves independently in practice, we could separately optimize each SKU to obtain the optimal discounts for all the SKUs. However, "global" business constraints tie different SKUs together. We introduce two kinds of global constraints in this paper. The first kind is from a business planning perspective, where each country has its own growth plans. For instance, the company may want the total sales for all the SKUs within a product category (e.g., women footwear) to reach a specific target. The second kind of business constraint comes from the cost perspective. Controlling the average discounts over a group of SKUs is critical to evaluating the cost spent on the campaigns. For example, the company may want to maintain the weighted average discount for a group of SKUs within a range (e.g.,15% to 17%). This value is denoted as the sales Discount Rate (sDR) target. The way Zalando adopts to

measure weighted average discount is to use discount rates weighted by their potential contributions to the total revenue (undiscounted price times sales), as defined below:

$$sDR_i = \frac{\sum_{(k,c,t)\in\mathcal{T}_i} \sum_l d_l \; x_{k,c,t,l} P_{k,c}}{\sum_{(k,c,t)\in\mathcal{T}_i} \sum_l x_{k,c,t,l} P_{k,c}}$$

sDR is critical from business planning perspective, as it measures the relative cost of adopting a discount strategy. The higher the sDR value, the more Zalando need to invest (or bear as opportunity costs) to the discount plan. In reality, sDR targets are closely monitored by Zalando high level business teams, and they usually impose certain values as targets to reach. For example, it might be required that for women footwear category, the weighted average discount (sDR) for Germany in the next four weeks is around 15%. We discuss in this section the global problem in which we jointly optimize a group of SKUs such that the above two types of constraints are satisfied in certain countries and weeks. We follow the notation in Section 4.4 and add subscript k to denote the associated variables of SKU k. Define target set $\mathcal{T}$ to be the set of SKUs, with specific sets of countries and time periods that we want to reach a certain global target. Given target set $\mathcal{T}_i := \{(k, c, t) \mid k \in \mathcal{K}_i, c \in \mathcal{C}_i, t \in \mathcal{W}_i\}$, we have two types of steering targets: revenue target and sDR target.

Revenue targets.

$$\sum_{(k,c,t)\in\mathcal{T}_i} \sum_l P_{k,c,t,l} x_{k,c,t,l} \geq GMV_i^- \tag{4.29}$$

sDR targets.

$$sDR_i^- \leq \frac{\sum_{(k,c,t)\in\mathcal{T}_i} \sum_l d_l \; x_{k,c,t,l} P_{k,c}}{\sum_{(k,c,t)\in\mathcal{T}_i} \sum_l x_{k,c,t,l} P_{k,c}} \leq sDR_i^+ \tag{4.30}$$

4.5.1 Lagrangian relaxation

The solutions of different SKUs are coupled via the global steering targets, and directly solving the global optimization problem for hundreds of thousands of SKUs could be an impossible task. Moreover, parallel computing power cannot be utilized if we optimize all the SKUs jointly. We, therefore, decompose the global target constraints by using Lagrangian relaxation, namely, by introducing dual multipliers, each

associated with a global target.

Let $\mathcal{P}_k$ be the feasible region of sub-problem k, which specifies the values of x_k and z_k subject to the constraints for single SKU optimization problem ((4.3)-(4.7), (4.9), (4.13)-(4.21), and (4.22)-(4.26)) for each single SKU k. We can then write the global problem as the following.

$$\max \sum_{k,c,t,l} \pi_{k,c,t,l} x_{k,c,t,l} + y_{k,end} \mathrm{SV}_k \tag{4.31}$$

$$s.t. \sum_{(k,c,t)\in\mathcal{T}_i} \sum_{l} P_{k,c,t,l} x_{k,c,t,l} \geq \mathrm{GMV}_i^- \quad \forall i \in I \tag{4.32}$$

$$\sum_{(k,c,t)\in\mathcal{T}_i} \sum_{l} (d_l - \mathrm{sDR}_i^-)\, x_{k,c,t,l} \geq 0 \quad \forall i \in I \tag{4.33}$$

$$\sum_{(k,c,t)\in\mathcal{T}_i} \sum_{l} (d_l - \mathrm{sDR}_i^+)\, x_{k,c,t,l} \leq 0 \quad \forall i \in I \tag{4.34}$$

$$x_k, z_k \in \mathcal{P}_k \tag{4.35}$$

Let $\lambda^-, \mu'^-, \mu'^+$ be the dual non-negative vectors associated with constraint (4.32), (4.33) and (4.34) respectively. Let $\theta = (\lambda, \mu^-, \mu_i^+)$. We then obtain the Lagrangian dual problem:

$$\min_{\theta\geq\mathbf{0}} g(\theta) := \max \sum_{k} \sum_{c,t,l} \Bigg(\pi_{k,c,t,l} + \sum_{\mathcal{T}_i} \lambda_i^- P_{k,c,t,l} + \sum_{\mathcal{T}_i} \mu_i^- (d_l - \mathrm{sDR}_i^-) \tag{4.36}$$

$$- \sum_{\mathcal{T}_i} \mu_i^+ (d_l - \mathrm{sDR}_i^+) \Bigg) x_{k,c,t,l} - \sum_{\mathcal{T}_i} \lambda_i \mathrm{GMV}_i^- + y_{k,end} \mathrm{SV}_k \tag{4.37}$$

$$s.t.\ \ x_k, z_k \in \mathcal{P}_k \tag{4.38}$$

where the Lagrangian dual function g is piece-wise linear, convex, continuous and non-smooth.

In general, Lagrangian multipliers θ must be non-negative to penalize target violations in the objective. In practice, the corresponding constraints could be infeasible, where the multipliers will go to infinity or be heavily violated in a given solution. In this case, the multipliers could have very large values, possibly orders of magnitude

larger than the rest of the objective function, causing slow convergence or numerical instability. To guarantee that the problem of optimizing $g(\theta)$ is both *bounded* and *numerically stable*, we assume each lagrangian multiplier θ_i to be bounded from above by a reference value $\bar{\theta}_i$ which can be easily computed from the data as follows. Consider a generic formulation for the ith target

$$\sum_{j \in J} \tau_{ji} x_j \leq T_i$$

with primal variables x_j, $j \in J$, and multiplier θ_i. Let $f(\mathbf{x}) = \sum_{j \in J} c_j x_j$ be the objective function to be maximized. Dualizing the ith target constraint yields the modified objective function

$$\min_{\theta_i} \left(\max_{\mathbf{x}} \sum_{j \in J} c_j - \theta_i \tau_{ji}) x_j \right) - \theta_i T_i$$

Then, the reference value for the multiplier has to satisfy the condition

$$\bar{\theta}_i > \bar{\theta}_i^{\min} = \max_{j \in J} \frac{c_j}{\tau_{ji}}$$

The condition guarantees that if target i is violated then the multiplier λ_i can become large enough to dominate the coefficients of the primal variables $\mathbf{x}$, i.e. $\max_{\theta_i \in [0, \bar{\theta}_i]} \theta_i \tau_{ji} > c_j \ \forall j \in J$ and thus steer the optimization towards target-reaching solutions. From a business perspective, the multiplier θ_i represents the per-sold-item cost of violating target j by an additional unit; hence the condition states that the maximum allowed cost per sale for an additional unit of violation $\bar{\theta}_i$ for any article $j \in J$ needs to be greater than the per-unit profit c_j for the same sale to guarantee convergence to a target-reaching solution. In our experiments we set $\bar{\theta}_i = 3\bar{\theta}_i^{\min}$ $\forall i \in I$.

4.5.2 Cutting plane algorithm

We use the Cutting Plane (CP) method (Kelley 1960) to solve (4.36). The method iteratively constructs a piece-wise linear approximation of the Lagrangian dual function g and minimizes its value, yielding a new dual vector of multipliers at each iteration. Specifically, given dual variables θ_0^n and $y_0 = g(\theta_0^n)$ at iteration n, we add the constraint or "(optimality) cut" $y \geq g'(\theta_0)(\theta - \theta_0) + y_0$ to the CP model:

$$\left(\mathcal{M}^n\right) \; z^n := \min \; y \tag{4.39}$$

$$s.t. \;\; y \geq g'(\theta_0^i)(\theta - \theta_0^i) + y_0^i \;\; \text{for} \;\; i = 1, \ldots, n \tag{4.40}$$

$$\mathbf{0} \leq \theta \leq \bar{\theta} \tag{4.41}$$

We chose the CP algorithm because it can provide good solution quality and rapid convergence while being relatively easy to implement in the company's environment and requiring little overhead compared to the solution of the lagrangian subproblems. See Frangioni et al. (2015) for a more in-detail evaluation of optimization methods for Lagrangian dual functions.

In the following, we report relevant properties of the CP algorithm. First, as g is convex, we have that the series $\{z_n\}_{n \in N}$ is non-decreasing, i.e. $z_n \geq z_{n-1} \; \forall n \in N$. Let S^* be the optimal value for the original problem and θ^* the optimal solution for the lagrangian dual problem. From strong-duality in convex problems, we will have $S^* = g(\theta^*)$, and $z^n \leq g(\theta^*) \leq g(\theta^n)$ for all iterations n, with the values converging to the optimum $g(\theta^*)$ in a finite number of steps. We can then define the optimality gap at iteration n as $\text{gap}_n = \dfrac{g(\theta^n) - z^n}{|z^n|}$. However, the MIP integer constraints introduce non-convexity to our problem, and we have instead the weak duality $S^* \leq g(\theta^*)$, and z^n converges to an upper bound of the optimal value of the original problem, i.e. $\exists \; \bar{n} \in N : z^n \geq S^n \; \forall n \geq \bar{n}$. Indeed, z^n converges to the value of the continuous relaxation of the equivalent Dantzig-Wolfe reformulation of the lagrangian relaxation (Desrosiers and Lübbecke 2005).

We report the algorithm scheme in figure 4. We initialize the algorithm with the "dummy solution" $\theta = \mathbf{0}$, which yields the "unconstrained optimum" or the profit-

maximal solution where targets are ignored. Other initialization schemes could yield better results or faster convergence depending on the problem and the underlying data. We refer to the literature Frangioni et al. (2015) for further information. In our experiments, the zero-multipliers solution is easier to use. They do not require any prior knowledge or computation and provide a benchmark of the optimization model without any global constraints.

As a stopping criterion, we could use a MIP optimality gap by computing at each iteration n a lower bound for the original problem using some heuristic, possibly based on the current lagrangian solution $(\theta_n, \mathbf{x}_n)$, and comparing it with the current best dual bound $\max_{n' \leq n} g(\theta'_n)$. Given the scale of our problem, computing accurate heuristic solutions during the iterations would be significantly expensive. We then consider the following stopping conditions: (i) the number of iterations has reached an upper limit N and (ii) the change between multipliers in subsequent iterations is below a minimum threshold ϵ.

Algorithm 4: Cutting Plane

Initialization. Optimization model $\mathcal{M}$; set $n = 1$, $z^0 = -\infty$; set dual values $\theta^0 = 0$
for $n = 1, \ldots, N$ **do**
  solve the Lagrangian relaxation problem to obtain $g(\theta^n)$;
  calculate subgradient $g'(\theta^n)$;
  if $n = N$ or $\|\theta^n - \theta^{n-1}\| > \epsilon$ **then**
    add optimality cut to model $\mathcal{M}^{n-1}$ yielding model $\mathcal{M}^n$
    solve model $\mathcal{M}^n$ to obtain dual value z^{n+1} and dual solution θ^{n+1}
  else
    break;
  end if
end for

4.5.3 Primal heuristics

While optimizing the dual function leads to finding primal solutions with minor violations and good objectives, the process does not guarantee to find primal solutions that are either optimal or feasible for the original problem. To tackle this challenge, we develop primal heuristics to construct a good global solution using all the results in the previous iterations.

Let $s_n = (p_n, v_{n1}, v_{n2}, \ldots, v_{n|J|})$ denote the result in iteration n where p_n is the profit objective and v_{nj} is the violation in global target j. Let $\bar{p} = \max_{n \in N}\{p_n\}$ be the highest profit in all iterations. Let σ_j be the target value of global target j. In addition, we record the profit and violations of each SKU k. Let p_{kn} be the profit from SKU k in iteration n and l_{knj} violation from SKU k for target j in iteration n. We then solve the following problem to obtain an optimal combination of the solutions.

$$\min \quad \sum_{j \in |J|} \delta_j + \delta_p \tag{4.42}$$

$$s.t. \quad \delta_p = 1 - \frac{\sum_{k \in K} \sum_{n \in N} r_{kn} p_{kn}}{\bar{p}} \tag{4.43}$$

$$\delta_j = \frac{\sum_{k \in \mathcal{T}_j} \sum_{n \in N} r_{kn} l_{knj}}{\sigma_j} \quad \forall j \in J \tag{4.44}$$

$$\sum_{n \in N} r_{kn} = 1 \tag{4.45}$$

$$r_{kn} \in \{0, 1\}; \ \delta_j, \delta_p \geq 0 \tag{4.46}$$

Figure 4-3 showcases one run of a global optimization problem with 1,000 SKUs and four sDR targets. It took eight iterations to converge within the optimality gap. The x-axis plots the percentage profit gap compared to the highest profit the model has ever seen. The y-axis measures the total amount of violations of the four sDR constraints. It is clear that when the algorithm stops at the eighth iteration, the primal solution is not particularly preferable. However, after computing a new solution with the Primal Heuristic, the violations are reduced to zero without compromising the profit objective.

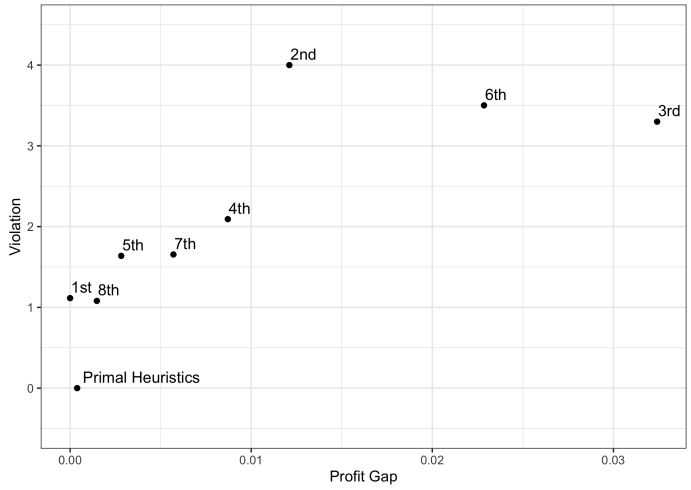

Figure 4-3: Profit Gaps and Violations Across Iterations and Primal Heuristic

4.6 Field Experiments

We have proposed various techniques in previous sections to model, solve and improve the discount optimization process. Zalando is very collaborative and eager to conduct offline and online experiments to validate the proposed framework. In this section, we report experiments performed at Zalando. These are the first stage of the experiments, with the primary purpose of checking the functionality of the system, as well as the ability to reach the global targets. In later field experiments, the main goal will shift towards measuring the improvement in total profits compared to the old system.

4.6.1 Offline Large-Scale Experiments

Before carrying out real field experiments, we first test the optimization framework offline by using historical data. From an experimental design perspective, we selected a sample of the assortment of the Kids category because it includes articles of all types, from shoes to accessories. Therefore its variety is comparable to the one of the whole shop. Other categories are split by gender or article type (Men Textile, Women Shoes, Accessories). In this sense, a sample from the Kids category is likely to be more representative of the whole shop than a sample of the same size from another category. The optimization spans 14 countries and 26 weeks. The goal is to

generate profit-optimal target-reaching discounts to be uploaded for the first week of the optimization horizon. The problem has one global sDR target for the first week of the experiments. To speed up the process and also facilitate the weekly optimization routine, the implementation exploits large-scale cloud-based parallelization to solve lagrangian subproblems in parallel. We use Amazon EMR as the execution platform, which allows the provision of Hadoop clusters with a specified amount of total cores and memory. Our implementation is written in Python, using the Apache Spark framework. In our experiments, we used the C5 instance type provided by AWS.

We present the results of two large-scale tests we run on our algorithm. The goal of these tests is to mimic the settings of the following field experiments, and the experiments started on different weeks. Run time is measured as the total time for the overall system to confirm the completion of each run after a launch. It can be seen that our algorithm manages to reach the targets in both cases.

Table 4.3: Offline experimental results

Experiments	SKUs	Iterations	Run Time (s)	Number of Cores	Total Memory	Target Deviation
1	51745	5	105	2000	4.0	0.00
2	12798	5	90	1500	3.0	0.00

4.6.2 Online Field Experiment Results

The online experiments were launched in consecutive weeks, where the model was solved using 2000 parallel processing cores and 4.0 TiB for each experiment. Bold values indicate when the model and the actual values in the target group satisfied the targets. It can be observed that the target (sDR) value for both the model and the test articles are close to or within the target bounds most of the time. For the test group, the actual sDR is much closer to the target than the control one, especially for experiment #3, proving the practical effect of the decisions taken by the model. More precisely, for test #3, the users later confirmed the targets were much harder than expected concerning the assortment, meaning our model managed to significantly

move the sDR for the test group despite the targets being difficult to reach. To compare profit, we require the Treatment and Control group to be subjected to the same (sDR) targets, as they directly impact profit.

During the field test, the Control group is managed directly by the commercial team using different heuristic processes that can yield different sDRs than the one for the Treatment group. Indeed, we observed large deviations in the Control group from the sDR target, primarily attributed to this heuristic process and other process-related issues. For this reason, we cannot directly compare profit among the two groups.

Table 4.4: Online field experimental results

Experiments	SKUs	Iterations	Run Time	Target deviation		
				Model	Actuals	
					Test	Control
1	12632	4	1h	-0.09%	**0.00%**	-3.27%
2	12757	1	40m	**0.00%**	**0.00%**	0.00%
3	8961	4	1h	-0.14%	-2.85%	-10.43%

4.7 Aggregation Model

In reality, the company is dealing with a global problem with potentially $K > 600,000$ SKUs, $C = 17$ countries, $T = 40$ weeks and $L = 15$ discount levels. Moreover, the global optimization in previous sections may take several iterations to terminate. Practically it imposes heavy burdens on computational resources and demands simplification or certain levels of aggregation. For the global problem, we propose an aggregation framework to cluster similar SKUs within each category into dummy SKUs. This framework brings value from several perspectives. Firstly it could significantly reduce the number of SKUs in the global problem, saving much computational time and resources. Secondly, there are many "long tail" SKUs that are less popular and have few or no historical sales. Since these SKUs are expected to have lower demand prediction accuracy and lower sales, therefore are less important to the business operations. By clustering, we combine these SKUs and make centralized price

decisions. Finally, aggregation results are also helpful to business planning. The aggregated output provides quick suggestions for business users on their high-level promotional planning for certain countries or product categories.

4.7.1 Clustering

On the global optimization level, the company has country-specific and category-specific targets. For instance, Figure 4-4 illustrates an example of a global optimization with 14 country targets (as indicated for each column) and four category targets (as indicated for each row). The values in individual cells are category-country-specific targets and are unknown before solving the global optimization problem. The idea of the aggregation model is to approximate the global problem by clustering SKUs within each category. The resulting model is much smaller due to clustering, whose outputs approximate the cell category-country targets. Finally, we could decompose the global problem into category-specific problems and solve each SKU's optimal discounts by using cell-specific targets.

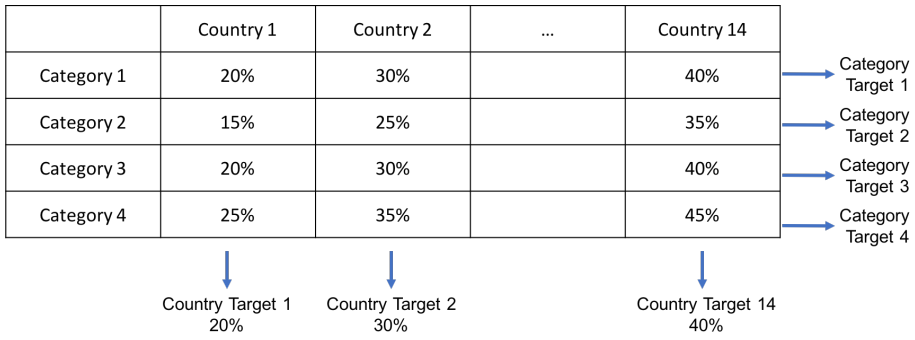

Figure 4-4: Example of Country and Category Global Targets

The first part of the aggregation model is to find a clustering algorithm. Ideally, the SKUs within each cluster should have similar optimal discounts since we combine their information. After exploring the features that impact optimal discount decisions, we include demand, price, inventory, and unit profit as features to perform clustering. After removing the outliers (SKUs with extremely low inventories or demands), we normalize the values on each dimension and apply the K-Means clustering algorithm

for each category.

4.7.2 Aggregation Approximation

After we form clusters in each product category, we need to apply an aggregation method to represent SKUs within each cluster. We consider two approaches. The first approach is to aggregate the SKUs in each cluster by taking the N closest SKUs to the cluster's center and scaling up demand, inventory, and other parameters accordingly. This approach is simple to implement; however, it only utilizes the information of a small number of SKUs and will not yield a good approximation if the cluster is spread out. The second approach is to aggregate the SKUs in each cluster into a giant dummy SKU. The challenge of this method is how to combine the inputs from different SKUs. We adopt the second approach and take the demand-weighted average of price, inventory, and other parameters.

4.7.3 Experimental Results

The data set contains 1,000 SKUs from 11 categories across 14 countries and 40 weeks of the planning horizon. The business users impose four global sDR targets, two on individual product categories (Women Textile and Women Footwear) and two on respective countries.

We test the aggregation framework by comparing it to the benchmark instance, where we apply the dis-aggregated method and jointly optimize SKUs from all the categories. In both cases, we implement methods for global optimization in Section 4.5, namely Lagrangian relaxations and primal heuristics on the MIP. We compare the running time and sDR targets for the two instances.

After solving the first stage of the aggregated model, value has already been created in category-level business insights. If needed or further required by the business users, we could solve each category on the dis-aggregated model, using sDR inputs from the aggregated model. The optimal objective difference for both product categories is less than 0.5%. We present the discount distribution histogram in Figure 4-5

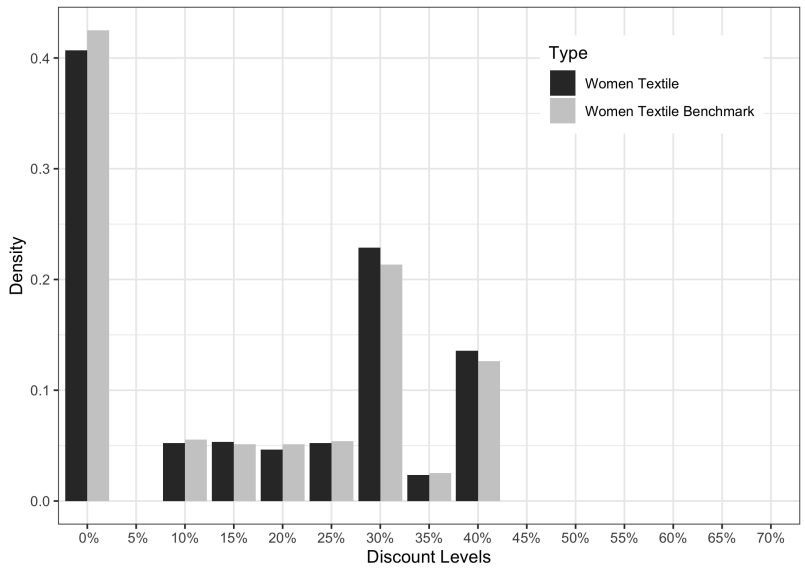

Figure 4-5: Discount Distribution Comparison for Women Textile Category

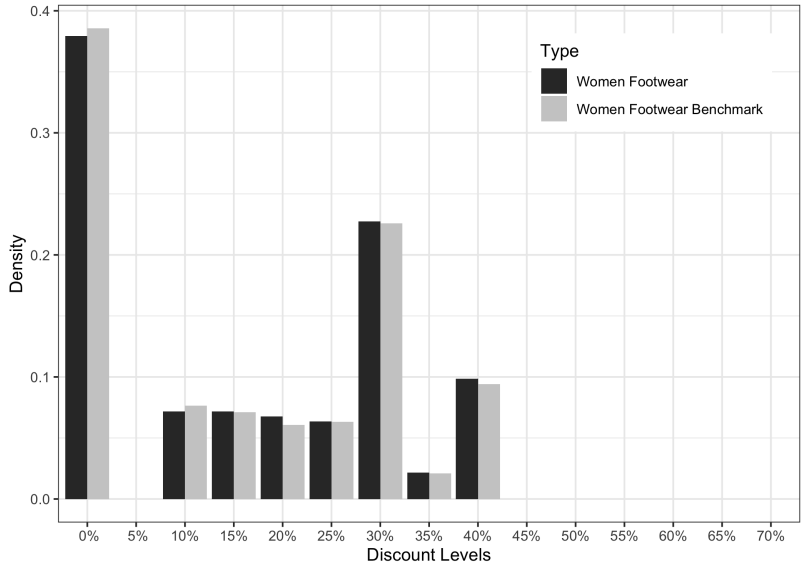

Figure 4-6: Discount Distribution Comparison for Women Footwear Category

Table 4.5: Comparison of sDR targets

	Country 1			Country 2		
	Benchmark	Aggregation	Difference	Benchmark	Aggregation	Difference
Women Textile	0.138	0.129	-0.009	0.161	0.171	0.01
Women Footwear	0.148	0.177	0.029	0.183	0.181	-0.002

and 4-6.

Chapter 5

Concluding Remarks

This chapter summarizes the previous works and discusses future research directions.

5.1 Assortment Display, Price Competition, and Fairness in Online Marketplaces

This paper investigates how platforms such as Airbnb should display their assortment to maximize total revenue. Specifically, we consider a stylized model where the profit-maximizing platform determines the partition of the products and traffic assigned to each partition. Each seller on the platform supplies a distinct product with one-unit inventory and sets the product's price given the prices of the other sellers in the partition. We provide theoretical justification that it is optimal to display the entire product assortment when a platform faces sufficiently large demand. Moreover, to derive the equilibrium price for each seller and ultimately offer recommendations on the display policy, we tabulate the FOC and formulate the platform's problem as an MIP. Exploiting the MIP, we can compute the equilibrium pricing for each seller and, based on this, recommend the corresponding optimal display policy for the platform. Our theoretical and numerical results carry critical managerial implications. The platform should adopt different display strategies to maximize its revenue depending on the market conditions characterized by the number of sellers

and customers. Additionally, to address the concerns over unfairness in the partition display from both the sellers and customers, we incorporate the fairness constraints on the closeness of (1) attractiveness of each partition, (2) the demand allocated to each partition, and (3) the individual customer welfare. Such fairness constraints guarantee a certain degree of fairness for the sellers and customers. Furthermore, we also use data from Airbnb to demonstrate how our framework can be applied in reality and simultaneously demonstrate the revenue loss that Airbnb incurs to achieve a fair display. Finally, we extend our current model to the case in which each seller has more than one unit of the listed product and discuss the optimal display policy when demand is sufficiently high.

Our MIP framework can generally be applied to platforms on which the service provider supplies a single unit of product. Such cases include Airbnb and the labor market, where each company posts a job with one vacancy and competes with other companies on salary, and eBay, where each seller posts one product and competes on price through the “buy it now" option. Nevertheless, our MIP framework is also flexible. It can include other industry-specific constraints, but importantly, it can also be applied to characterize the equilibrium price and derive the optimal display policy for platforms on which sellers have more than one unit of inventory in stock.

5.2 Estimating and Exploiting the Impact of Photo Layout: A Structural Approach

This paper studies how the photo layout on Airbnb's listing web pages affects customers' renting decisions. Specifically, we provide an integrated framework to extract photo features, evaluate photo impact and ultimately optimize the photo layout. This framework can be adopted by platforms such as Yelp and Hotel.com to help select and recommend user-generated images in the user comments. Similar to the case of Airbnb, such images tend to vary significantly in quality and content. Our framework can be used to estimate and optimize the image recommendation system to boost the

user engagement level. Moreover, online retailers, such as Farfetch and Alibaba, can use our framework to identify photo content and photo features that boost demand. According to Martinez de Albeniz (2017), the image production process for online retailers is usually time-consuming and requires careful planning on the operations side. By providing guidance on the photo layout and the photo features to be included, our framework can improve the conversion rate for the platform and potentially also simplify the image production process.

Specifically, we take advantage of the advancement in computer vision algorithm to augment the Airbnb transaction datasets with information about the photo quality and room of images posted by listing owners. Then, to address two estimation challenges associated with the Airbnb setting, namely shrinking consideration set and censored demand, we develop a pairwise comparison model (PCM) that utilizes information about the sequencing of property bookings to supplement the sales data. The PCM estimation framework allows for flexible utility structure and is asymptotically normal with a convergence rate linear in the number of observations. We find that the cover photo generates a much higher impact than the rest of the photos, possibly because (i) the cover image is displayed when customers are browsing through different listing options on the search page and (ii) the cover image takes up a more significant amount of space on the listing's web page than do the other photos. Additionally, the estimation results suggest that a high-quality bedroom photo is the best candidate for the cover image of an Airbnb listing. We determine the optimal photo layout for each listing based on our estimation results. Our counterfactual analysis suggests that a listing's unilateral adoption of optimal photo layout leads to 11.0% more bookings on average, which translates into an average annual revenue improvement of $1248. Moreover, depending on the location and market size, when all listings simultaneously switch to the optimal photo layout, they get booked for two to five additional days on average, which is equivalent to $500 to $1100 increase in revenue, respectively.

The PCM estimation framework can be applied practically to settings where products have limited quantities, and the sequence of purchase is observed, e.g., the car rental market, second-hand market, and flash sale websites. As each SKU has lim-

ited units in stock in such cases, the aggregate demand data are usually censored, and customers' consideration set changes frequently. PCM is particularly suitable for such scenarios, as it can uncover the true values for the parameters of interest under changing consideration sets even when the no-purchase data are *not* observed.

Our study empirically quantifies the impact of photo layout and contributes to understanding the impacts of visual messages in the online retailing and advertisement industries. There are a few limitations and directions for future work. First, with more detailed data on customers' devices and clicking and searching behaviors, we will be able to focus on web users and better construct their consideration sets, thus obtaining more refined estimation results. Additionally, extending the current layout definition with a larger image dataset would be interesting, and exploring the impact of the display sequence of photos in the non-cover spot on listing demand. Given sufficient variation in the room type of the first five images, the impact of the photo layout under alternative definitions of duplication can also be tested. Meanwhile, the estimation results can be verified through a lab experiment in which each listing has infinite availability each day. The estimation process is simplified, as it can be carried out using the standard MNL model. Finally, the photo layout optimization problem can be studied in conjunction with product assortment optimization or price optimization to generate a more substantial revenue increase for the platform.

5.3 Large-scale Price Optimization for an Online Fashion Retailer

Zalando has successfully implemented all the algorithms described in this work into their weekly price discount recommendation system routine. The optimization model outputs optimal discount decisions for hundreds of thousands of products every week across more than 14 markets. As the preliminary field experiments suggest, the model can steer the discounts to satisfy the global targets. In conclusion, our work addresses several critical challenges for the online fashion retailer Zalando, with a vast number

of products and different levels of business constraints. We first manage to model the single product optimization problem as an MIP instance, which correctly captures the challenge of stock hedging and stock response. On the “global" level with constraints across multiple products, we apply the Lagrangian decomposition and the cutting plane method to efficiently find the solution within the optimality gap for this large-scale optimization problem. We also adopt a heuristic to combine solutions across iterations to yield a better solution. Finally, we propose an aggregation framework that will drastically improve the computational time and provide high-level business insights. The pilot field experiment empirically validates that the optimization framework successfully steers the discounts towards the business targets, and the model will be integrated into the company's weekly operation pipeline. We provide a new algorithm system that automates the decision-making process for a globally operating fashion platform. This system can be applied to many other similar business environments in the future. We plan to test the algorithms on the company platform and investigate the profit improvement under this new decision-making system to show more convincing results. Another important future direction is how to reduce the scale of the problem or increase the efficiency of our system so we can achieve better speed without losing much optimality.

Appendix A

Proofs and Alternative Formulations for Chapter 2

A.1 Proofs for Propositions and Theorems

A.1.1 Proof for Proposition 2.1

For existence, we can show the derivative of revenue for listing i with respect to p_i is

$$\frac{d\Pi_i}{dp_i} = 1 - (M\beta p_i q_i + 1)(1 - q_i)^M. \tag{A.1}$$

Clearly when $p_i = 0$, we have $\frac{d\Pi_i}{dp_i} = 1 - (1 - q_i)^M > 0$, and when p_i is very large, we can show that $\frac{d\Pi_i}{dp_i} \to 0$ and $\frac{d\Pi_i}{dp_i} < 0$. This is because as $p_i \to \infty$ we have $q_i \to 0$ and $p_i q_i = \frac{p_i}{1+S_{-i}e^{\beta p_i - a_i}} \to 0$, where $S_{-i} = 1 + \sum_{j \neq i} e^{a_j - \beta p_j}$ is a constant, so $\frac{d\Pi_i}{dp_i} \to 0$ and

$$\lim_{p_i \to \infty} \frac{M \ln(1 - q_i) + \ln(M\beta p_i q_i + 1)}{M\beta p_i q_i} = \lim_{p_i \to \infty} \frac{\ln(1 - q_i)}{\beta p_i q_i} + \frac{\ln(M\beta p_i q_i + 1)}{M\beta p_i q_i} = 1 > 0.$$

This implies $(M\beta p_i q_i + 1)(1 - q_i)^M > 1$ when p_i is large and we get $\frac{d\Pi_i}{dp_i} < 0$ immediately.

Now consider the second order derivative

$$\frac{d^2\Pi_i}{dp_i^2} = -(1-q_i)^M(M\beta q_i - \beta^2 M p_i q_i(1-q_i)) - M\beta q_i(1-q_i)(1-q_i)^{M-1}(M\beta p_i q_i \quad \text{(A.2)}$$

$$= -M\beta q_i(1-q_i)^M(M\beta p_i q_i - \beta p_i(1-q_i)+2). \quad \text{(A.3)}$$

When $p_i = 0$, we have $\frac{d^2\Pi_i}{dp_i^2} = -2\beta M q_i(1-q_i)^M < 0$. From the expression above, we can see directly that $M\beta p_i q_i - \beta p_i(1-q_i)+2 \to -\infty$ and $q_i(M\beta p_i q_i - \beta p_i(1-q_i)+2) \to 0$. This means that when p_i is large, $\frac{d^2\Pi_i}{dp_i^2} > 0$ and approaches to zero. Because the expression is a continuous function of p_i and flips the sign on the domain $[0, +\infty)$, we know there exists at least one root. Setting the second order derivative to zero, we get $Mq_i + q_i - 1 = -\frac{2}{\beta p_i}$. The left hand side is a decreasing function in p_i and the right hand side is increasing in p_i, and there can be at most one intersect. Now we know that there exist a single non-saddle p^{**} such that $\frac{d^2\Pi_i}{dp_i^2} = 0$. It is then obvious that $\frac{d\Pi_i}{dp_i}$ is decreasing on $[0, p^{**}]$ and increasing on $[p^{**}, +\infty)$. Knowing that $\frac{d\Pi_i}{dp_i}$ is positive when $p_i = 0$, and it approaches to zero from the negative side when p_i is large, we conclude that there exists a single root p^* for the first derivative, so that $\frac{d\Pi_i}{dp_i} > 0$ on $[0, p^*]$ and $\frac{d\Pi_i}{dp_i} < 0$ on $[p^*, +\infty)$. This proves quasi-concavity of Π_i as a function of p_i.

Since $q_i = \frac{e^{a_i - \beta p_i}}{1+\sum_j e^{a_j - \beta p_j}} = \frac{1}{1+S_{-i}e^{\beta p_i - a_i}}$, where $S_{-i} = 1 + \sum_{j\neq i} e^{a_j - \beta p_j} \in (1, 1 + \sum_j e^{a_j}) \triangleq (1, S)$ is bounded, we have

$$(M\beta p_i q_i + 1)(1-q_i)^M > (M\beta p_i \frac{1}{1+Se^{\beta p_i - a_i}} + 1)(1 - \frac{1}{1+e^{\beta p_i - a_i}})^M \triangleq f_i(p_i).$$

Following the same argument before, we can prove the existence of p_i^* such that $f_i(p_i) > 1$ when $p_i > p_i^*$. This means $\frac{d\Pi_i}{dp_i} < 0$ when $p_i > p_i^*$. Note that p_i^* only depend on the qualities $\{a_j\}_{j=1}^n$, so we can restrict the price space to be $\Pi_i[0, p_i^*]$, which is a compact set. Fudenberg and Tirole (1991) implies existence of pure strategy Nash equilibrium under such condition.

For uniqueness, let $H = \{\frac{\partial^2\Pi_i}{\partial p_i \partial p_j}\}_{N\times N}$, from Cachon and Netessine (2006) we only need to prove that $(-1)^N|H|$ is positive at the equilibrium. From direct calculation,

we get the following

$$\begin{aligned}\frac{\partial^2\Pi_i}{\partial p_i^2} &= -M\beta q_i(1-q_i)^M(M\beta p_i q_i - \beta p_i(1-q_i)+2)\\ \frac{\partial^2\Pi_i}{\partial p_i \partial p_j} &= M\beta q_i q_j(1-q_i)^{M-1}(M\beta p_i q_i - \beta p_i(1-q_i)+1).\end{aligned}$$

By applying the calculation rules for determinant, the sign of $|H|$ is the same as the sign of $|A|$ where

$$A_{ii} = -(\frac{1}{q_i}-1)(M\beta p_i q_i - \beta p_i(1-q_i)+2), \qquad A_{ij} = M\beta p_i q_i - \beta p_i(1-q_i)+1.$$

Denote $h_i = M\beta p_i q_i - \beta p_i(1-q_i)+1$ and $r_i = 1 - \frac{h_i+1}{q_i}$, we have

$$|A| = \begin{vmatrix} h_1+r_1 & h_1 & \dots & h_1 \\ h_2 & h_2+r_2 & \dots & h_2 \\ \vdots & \vdots & & \vdots \\ h_N & h_N & \cdots & h_N+r_N \end{vmatrix} = r_1 r_2 \dots r_N \left(1 + \frac{h_1}{r_1} + \frac{h_2}{r_2} + \cdots + \frac{h_N}{r_N}\right).$$

We next prove $r_i < 0$ at equilibrium point. Indeed, $r_i < 0 \iff h_i + 1 - q_i > 0 \iff M\beta p_i q_i + 1 - \beta p_i(1-q_i) + 1 - q_i > 0$. From the FOC condition 2.2, we get $\beta p_i = ((1-q_i)^{-M}-1)/(Mq_i)$, plug this into the expression of r_i, we get

$$\begin{aligned} r_i < 0 &\iff (1-q_i)^{-M} - \frac{(1-q_i)^{-M}-1}{Mq_i}(1-q_i) + (1-q_i) > 0\\ &\iff (M+1)q_i + (Mq_i+1)(1-q_i)^{M+1} > 1\end{aligned}$$

By Bernoulli inequality and the fact that $M \geq 1$, we get

$$\begin{aligned}\text{LHS of A.1.1} &\geq (M+1)q_i + (q_i+1)(1-q_i)^{M+1}\\ &= (M+1)q_i + (1-q_i^2)(1-q_i)^M\\ &> (M+1)q_i + (1-q_i^2)(1-Mq_i)\\ &= 1 + (q_i - q_i^2) + Mq_i^3 > 1\end{aligned}$$

which proves $r_i < 0$.

Note that we have $h_i < 1 - q_i + h_i = -q_i r_i$, which implies $\frac{h_i}{r_i} > -q_i$ when $r_i < 0$. In this case, we can obtain $1 + \frac{h_1}{r_1} + \frac{h_2}{r_2} + \cdots + \frac{h_N}{r_N} > 1 - q_1 - q_2 - \cdots - q_N > 0$. So the sign of $|A|$ is $(-1)^N$ and we have $(-1)^N|H|$ is positive at the equilibrium point. This combines with the fact that Π_i is quasi-concave implies the uniqueness of pure strategy Nash equilibrium by Cachon and Netessine (2006).

A.1.2 Proof for Proposition 2.2

By applying Implicit Function Theorem to the i-th FOC (2.2) we obtain

$$\begin{aligned}\frac{\partial FOC_i}{\partial a_i} &= -(1-q_i)^M(M\beta p_i q_i(1-q_i)) - Mq_i(1-q_i)(1-q_i)^{(M-1)}(M\beta p_i q_i + 1) \\ &= -Mq_i(1-q_i)^M(M\beta p_i q_i - \beta p_i(1-q_i) + 1)\end{aligned}$$

We already have the partial derivative with respect to p_i from Equation (A.2). Hence we are able to calculate $\frac{\partial p_i}{\partial a_i}$ as

$$\frac{\partial p_i}{\partial a_i} = -\frac{\partial FOC_i/\partial a_i}{\partial FOC_i/\partial p_i} = \frac{M\beta p_i q_i - \beta p_i(1-q_i) + 1}{(M\beta p_i q_i - \beta p_i(1-q_i) + 2)\beta}, \tag{A.4}$$

Note that by applying the Bernoulli inequality to the FOC equation (2.2) , we can get

$$1 = (1 + M\beta p_i q_i)(1-q_i)^M > (1 + M\beta p_i q_i)(1 - Mq_i) \Longrightarrow 1 + M\beta p_i q_i > \beta p_i. \tag{A.5}$$

Hence we can conclude that $\frac{\partial p_i}{\partial a_i} > 0$. Furthermore, when $M \to \infty$, we can have $Mq_i \to \infty$ (this is a direct result from Lemma 2.1), so the right hand side of (A.4) approaches to $\frac{1}{\beta}$ and we finish our proof.

A.1.3 Proof for Lemma 2.1

For the first part of the lemma, if there exists p_U such that $0 < p_i < p_U$ for infinite many M, then there exists a lower bound q_L for q_i because $q_i = \frac{\exp(a_i - \beta p_i)}{1 + \sum_j \exp(a_j - \beta p_j)} =$

$\frac{1}{\exp(\beta p_i - a_i) + \sum_j \exp(a_j - a_i - \beta(p_j - p_i))} > \frac{1}{\exp(\beta p_U - a_i) + \sum_j \exp(a_j - a_i + 2\beta p_U)}$ and this expression does not contain p_i. Then consider the FOC equation, we have

$$-M\ln(1-q_L) < -M\ln(1-q) = \ln(M\beta pq + 1) < \ln(M\beta p_U + 1). \tag{A.6}$$

Take $M \to \infty$ and we immediately get contradiction since the LHS of (A.6) is linear in M while the order of the RHS of (A.6) is logarithm in M. We can also have $q_i \to 0$ by the MNL equation since we have now already proven all $p_i \to \infty$.

For the second part, we first prove $M\beta p_i q_i \to \infty$ as $M \to \infty$. Since $q_i \to 0$, from the FOC equation, we have:

$$\ln(M\beta p_i q_i + 1) = -M\ln(1-q_i) = Mq_i + o(Mq_i) \tag{A.7}$$

If there exists U^* such that $M\beta p_i q_i < U^*$ for infinite many M, then because $p_i \to \infty$, we must have $Mq_i \to 0$. This means the right side of (A.7) goes to 0, hence we get $M\beta p_i q_i \to 0$. However, in this case we can get $\ln(M\beta p_i q_i + 1) \sim M\beta p_i q_i$, which is a contradiction to (A.7) because we would have $M\beta p_i q_i \sim Mq_i$ and this cannot be true since $p_i \to \infty$.

Then we prove the second statement in the lemma. Note that from the MNL model, we have $\beta p_i = a_i - \ln q_i + \ln(\frac{1-q_i}{1+S_{-i}})$ where $S_{-i} = \sum_{j \neq i} e^{a_j - \beta p_j}$. Since $p_j \to \infty$, we have $S_{-i} \to 0$ and $\ln(\frac{1-q_i}{1+S_{-i}}) \to 0$. Because $\lim_{M\to\infty} M\beta p_i q_i = \infty$, plug this into the FOC, we get

$$\ln Mq_i + \ln(a_i - \ln q_i + \ln(\frac{1-q_i}{1+S_{-i}})) + o(1) = -M\ln(1-q_i)$$

Since $q_i \to 0$ as $M \to \infty$, we further get $\ln(a_i - \ln q_i + \ln(\frac{1-q_i}{1+S_{-i}})) \to \infty$ and $\lim_{M\to\infty}(\ln(a_i - \ln q_i + \ln(\frac{1-q_i}{1+S_{-i}})) - \ln\ln\frac{1}{q_i}) = 0$, so we have

$$\ln Mq_i + \ln\ln\frac{1}{q_i} + o(1) = -M\ln(1-q_i) = Mq_i + o(Mq_i).$$

Note that $\ln\ln\frac{1}{q_i} \to \infty$, so $Mq_i \to \infty$. In this case, we can have the order $\ln Mq_i =$

$o(Mq_i)$. By merging the term $\ln Mq_i$ to $o(Mq_i)$ we finally achieve $Mq_i \sim \ln\ln\frac{1}{q_i}$ as desired.

A.1.4 Proof for Lemma 2.2

Denote $q_1 = q_i^*(M, \mathcal{S}_1), q_2 = q_i^*(M/\gamma, \mathcal{S}_2)$, which is the probability of item a_i in the two sub-markets (we drop index i for simplicity). Then from Lemma 2.1 we have: $Mq_1 \sim \ln\ln\frac{1}{q_1}, \frac{M}{\gamma}q_2 \sim \ln\ln\frac{1}{q_2}$. Take the ratio of these two equations, we get

$$\lim_{M\to\infty} \frac{\gamma q_1}{q_2}\frac{\ln(-\ln q_2)}{\ln(-\ln q_1)} = 1 \tag{A.8}$$

We next prove by contradiction that the above equation (A.8) implies

$$\lim_{M\to\infty} \frac{\gamma q_1}{q_2} = 1.$$

Let $x = \ln\gamma q_1, y = \ln q_2$, then by taking logarithm of both sides of (A.8), we have

$$\lim_{M\to\infty} x - y + \ln\ln(-y) - \ln\ln(\ln\gamma - x) = 0$$

If $\lim_{M\to\infty} x - y \neq 0$, then without loss of generality, suppose there exists $\epsilon > 0$ such that $x - y > \epsilon$ for infinite many times. Take M large enough, we can require x, y to satisfy

$$x - y + \ln\ln(-y) - \ln\ln(\ln\gamma - x) < \epsilon/2$$

This means $\ln\ln(-y) - \ln\ln(\ln\gamma - x) < -\epsilon/2$, which implies $\ln(-y)/\ln(\ln\gamma - x) < e^{-\epsilon/2}$. Take exponential and we get

$$\epsilon - x < -y < (\ln\gamma - x)^{e^{-\frac{\epsilon}{2}}}.$$

This cannot hold when M is large enough because $e^{-\epsilon/2} < 1$ and we have $x \to -\infty$

when $M \to \infty$. As a result, we can conclude that

$$\lim_{M\to\infty} x - y = \lim_{M\to\infty} \ln(\frac{\gamma q_1}{q_2}) = 0,$$

which gives us $\lim_{M\to\infty} \frac{\gamma q_1}{q_2} = 1$ as desired.

Finally, we consider the revenue Π_i. From Lemma 2.1, we have $\lim_{M\to\infty} Mq_1 = \lim_{M\to\infty}(M/\gamma)q_2 = \infty$, so we are able to calculate the revenue difference as follows,

$$\begin{aligned}
\lim_{M\to\infty} \Pi_i(\mathcal{S}_1, M) - \Pi_i(\mathcal{S}_2, M/\gamma) &= \lim_{M\to\infty} (p_1 - p_2) - (\frac{p_1}{M\beta p_1 q_1 + 1} - \frac{p_2}{\frac{M}{\gamma}\beta p_2 q_2 + 1}) \\
&= \lim_{M\to\infty} p_1 - p_2 \\
&= \lim_{M\to\infty} \ln(\frac{q_2}{q_1}) + \ln(\frac{1-q_1}{1+S_1}) - \ln(\frac{1-q_2}{1+S_2}) \qquad \text{(By the FOC)} \\
&= \lim_{M\to\infty} \ln(\frac{q_2}{q_1}) \\
&= \ln\gamma.
\end{aligned}$$

A.1.5 Proof for Lemma 2.3

We first prove p_i is monotone increasing with respect to M, i.e., when the demand expands, the seller should charge higher price. We prove this by decoupling demand M and total attractiveness z. Specifically, by writing the FOC equation $(1+M\beta p_i q_i)(1-q_i)^M = 1$ as

$$\left(1 + M\beta p_i \frac{\exp(a_i - \beta p_i)}{1+z}\right)\left(1 - \frac{\exp(a_i - \beta p_i)}{1+z}\right)^M = 1, \tag{A.9}$$

we can define $p_i(z, M)$ to be the solution of the above equation (A.9). In this way, the real attractiveness z^* would be the solution of $\sum_{i=1}^N \exp(a_i - \beta p_i(z, M)) = z$. For any $M_1 < M_2$, suppose z_1 and z_2 to be the total attractiveness at equilibrium under demand M_1 and M_2, i.e., $\sum_{i=1}^N \exp(a_i - \beta p_i(z_1, M_1)) = z_1$ and $\sum_{i=1}^N \exp(a_i - \beta p_i(z_2, M_2)) = z_2$, then we only need to prove $p_i(z_1, M_1) < p_i(z_2, M_2)$.

By implicit function theorem, we have

$$\frac{\partial p_i(z,M)}{\partial z} = \frac{\beta p_i(1-q_i) - (1+M\beta p_i q_i)}{\beta(1+z)\left((1-\beta p_i)(1-q_i)+1+M\beta p_i q_i\right)},$$
$$\frac{\partial p_i(z,M)}{\partial M} = \frac{(\beta p_i q_i + \ln(1-q_i)(1+M\beta p_i q_i))(1-q_i)}{M\beta q_i((\beta p_i - 1)(1-q_i) - (1+M\beta p_i q_i))}.$$

Note that we already prove in (A.5) that $M\beta p_i q_i + 1 > \beta p_i$. Also we have $\ln(1-q_i) < -q_i$ by simple calculus. As a result, we obtain $\frac{\partial p_i}{\partial M} > 0$ and $\frac{\partial p_i}{\partial z} < 0$. Hence, we have $p_i(z_1, M_1) < p_i(z_1, M_2)$ as $M_1 < M_2$. We next prove $z_1 > z_2$. If we do have $z_1 > z_2$, this would result in $p_i(z_1, M_2) < p_i(z_2, M_2)$ and we have our desired result. In fact, let $f(z) = \sum_{i=1}^{N} \exp(a_i - \beta p_i(z, M_2)) - z$, then z_2 is the root of $f(z)$. Define $q_i(z, M) = \exp(a_i - \beta p_i)/(1+z)$, then we can calculate its derivative with respect to z as

$$\frac{\partial q_i(z,M)}{\partial z} = -\frac{q_i}{1+z} \cdot \frac{1-q_i}{((1-\beta p_i)(1-q_i)+1+M\beta p_i q_i)}. \tag{A.10}$$

For attractiveness level z_1, we have $q_i(z_1, M_2) < q_i(z_1, M_1) < 1$ because p_i is monotone increasing with respect to M. Also note that $q_i(z, M_2) \equiv 0$ and $q_i(z, M_2) \equiv 1$ is two solutions to the above ODE (A.10). Then by classical ODE theory, because $0 < q_i(z_1, M_2) < 1$, then $0 < q_i(z, M_2) < 1$ for any $z > 0$. As a result, we would have $\frac{\partial q_i(z,M)}{\partial z} < 0$. Hence, for any $z > z_1$, we can get $q_i(z, M_2) < q_i(z_1, M_2) < q_i(z_1, M_1)$ and we can calculate $f'(z)$ as

$$\begin{aligned}
f'(z) &= -\sum_{i=1}^{N} \exp(a_i - \beta p_i)\beta \frac{\partial p_i}{\partial z} - 1 \\
&= -\sum_{i=1}^{N} (1+z) q_i \beta \frac{\beta p_i(1-q_i) - (1+M\beta p_i q_i)}{\beta(1+z)\left((1-\beta p_i)(1-q_i)+1+M\beta p_i q_i\right)} - 1 \\
&= \sum_{i=1}^{N} q_i \frac{(1+M\beta p_i q_i) - \beta p_i(1-q_i)}{(1+z)\left((1-\beta p_i)(1-q_i)+1+M\beta p_i q_i\right)} - 1 \\
&\leq \sum_{i=1}^{N} q_i - 1 \\
&< \sum_{i=1}^{N} q_i(z_1, M_1) - 1 < 0.
\end{aligned}$$

Because we already have $f(z_1) = \sum_{i=1}^{N} \exp(a_i - \beta p_i(z_1, M_2)) - z_1 < \sum_{i=1}^{N} \exp(a_i - \beta p_i(z_1, M_1)) - z_1 = 0$, hence $f(z) < 0$ for every $z > z_1$ and we must have $z_2 < z_1$ since z_2 is a root of $f(z)$. As a result, p_i is increasing in M.

Then calculate $\frac{\partial \Pi_i}{\partial M}$ without decoupling z and M. In this case, M would be the only variable and z is determined by the system of FOC equations. By taking derivative with respect to M at both sides of the MNL equation, we have (write $q_i' = \frac{\partial q_i}{\partial M}$ and $p_i' = \frac{\partial p_i}{\partial M}$):

$$q_i' = -q_i \beta p_i' + q_i (\sum_{j=1}^{N} q_j \beta p_j').$$

So we can further calculate $\frac{\partial \Pi_i}{\partial M}$ as

$$\begin{aligned}
\frac{\partial \Pi_i}{\partial M} &= \frac{\beta p_i}{(M\beta p_i q_i + 1)^2} \left(p_i q_i + (M\beta p_i q_i + 1 + q_i) M q_i p_i' + M p_i q_i'\right) \\
&= \frac{\beta p_i}{(M\beta p_i q_i + 1)^2} \left(p_i q_i + (M\beta p_i q_i + 1 + q_i) M q_i p_i' + M\beta p_i (-q_i p_i' + q_i (\sum_{j=1}^{N} q_j p_j'))\right) \\
&\geq \frac{\beta p_i}{(M\beta p_i q_i + 1)^2} \left(p_i q_i + (1 + q_i) M q_i p_i' + M\beta p_i q_i (\sum_{j=1}^{N} q_j p_j')\right) \\
&> 0.
\end{aligned}$$

The first inequality is because of (A.5) and the second inequality is because we already prove $p_i' > 0$.

A.1.6 Proof for Theorem 2.2

By similar argument in Lemma 2.1, we can still have $\lim_{M\to\infty} q_i = 0$ and $\lim_{M\to\infty} p_i = \infty$. We then prove that $Mq_i \to \infty$ when $M \to \infty$. If this is not true, there must exists $U > 0$ such that $Mq_i < U$ for infinite many M, with out loss of generality, we assume this is true for all M. Denote $A_j(M) = (W_i - j)\binom{M}{j} q_i^j (1 - q_i)^{M-j} (1 + M\beta p_i q_i - j\beta p_i)$,

then the FOC can be written as $\sum_{j=0}^{W_i-1} A_j(M) = W_i$, and we can have

$$\begin{aligned}\frac{A_j(M)}{A_{j-1}(M)} =& \frac{(W_i-j)\binom{M}{j}q_i^j(1-q_i)^{M-j}(1+M\beta p_i q_i - j\beta p_i)}{(W_i-j+1)\binom{M}{j-1}q_i^{j-1}(1-q_i)^{M-j+1}(1+M\beta p_i q_i-(j-1)\beta p_i)} \\ =& \frac{W_i-j}{W_i-j+1}\frac{\binom{M}{j}}{\binom{M}{j-1}}\frac{q_i}{1-q_i}\frac{1+M\beta p_i q_i - j\beta p_i}{1+M\beta p_i q_i-(j-1)\beta p_i}\end{aligned} \tag{A.11}$$

which is bounded under our hypothesis. (We use the fact that $\binom{M}{j}/\binom{M}{j-1} = O(M)$ when $M \to \infty$.) Note that $\sum_{j=0}^{W_i-1} A_j(M) = W_i$, so each term $A_j(M)$ is bounded, which is true for $j = 0$. Then by the same argument in Lemma 2.1, we can have $Mq_i \to \infty$, which is a contradiction.

Since we now $\lim_{M\to\infty} Mq_i = \infty$, we can deduce that $\frac{A_j(M)}{A_{j-1}(M)} \to \infty$ by (A.11). However, the FOC requires $\sum_{j=0}^{W_i-1} A_j(M) = W_i$, so we must have $\lim_{M\to\infty} A_{W_i-1}(M) = W_i$ and $\lim_{M\to\infty} A_j(M) = 0, \forall j < W_i - 1$. This implies

$$\binom{M}{W_i-1}q_i^{W_i-1}(1-q_i)^{M-W_i+1}(1+(Mq_i-W_i+1)\beta p_i) \to W_i, \quad (M\to\infty) \tag{A.12}$$

We can analyze the order of the left hand side of equation (A.12) as we did in Lemma 2.2:

$$\text{LHS of A.12} = O\left((Mq_i)^{W_i-1}(1-q_i)^{M-W_i+1}Mq_ip_i\right).$$

By taking logarithm, we can further have

$$(W_i-1)\ln(Mq_i) + (M-W_i+1)\ln(1-q_i) + \ln(Mq_i) + \ln(p_i) = O(1).$$

Because $\ln(1-q_i) = O(q_i)$, we then achieve

$$Mq_i + \ln(p_i) = Mq_i + \ln\left(a_i - \ln q_i + \ln\left(\frac{1-q_i}{1+S_{-i}}\right)\right) = O(1).$$

Hence, we conclude $Mq_i \sim \ln\ln\frac{1}{q_i}$ by the fact that $S_{-i} \to 0$ and $q_i \to 0$. Then it

follows from the proof of Lemma 2.2 that

$$\lim_{M\to\infty} p_i(\mathcal{S}_1, M) - p_i(\mathcal{S}_2, M/\gamma) = \lim_{M\to\infty} \ln\left(\frac{q_i(\mathcal{S}_1, M)}{q_i(S_2, M/\gamma)}\right) = \ln\gamma.$$

Note that the FOC requires

$$\sum_{j=0}^{W_i-1}(W_i - j)\binom{M}{j}q_i^j(1-q_i)^{M-j}(1+(Mq_i-j)\beta p_i) = W_i.$$

this combines with the fact that $Mq_i \to \infty$ give us

$$\lim_{M\to\infty}\sum_{j=0}^{W_i-1}(W_i - j)\binom{M}{j}q_i^j(1-q_i)^{M-j}p_i = 0.$$

So the difference in revenue can be calculated as

$$\lim_{M\to\infty}\Pi_i(\mathcal{S}_1, M) - \Pi_i(\mathcal{S}_2, M/\gamma) = \lim_{M\to\infty} W_i(p_i(\mathcal{S}_1, M) - p_i(\mathcal{S}_2, M/\gamma)) = W_i\ln\gamma.$$

From here, the rest of the proof is similar to Theorem 2.1.

A.1.7 Proof for Proposition 2.3

We will prove this bound by using the decoupling technique as we did in the proof of Lemma (2.3). This means we will consider the total attractiveness and demand separately. In fact, by writing the FOC equation $(1+M\beta p_i q_i)(1-q_i)^M = 1$ as

$$\left(1 + M\beta p_i\frac{\exp(a_i-\beta p_i)}{1+z}\right)\left(1-\frac{\exp(a_i-\beta p_i)}{1+z}\right)^M = 1, \tag{A.13}$$

we can define $p_i(z, M)$ to be the solution of the above equation (A.13) and define the corresponding $q_i(z, M) = \exp(a_i - \beta p_i(z, M))/(1+z)$ and $\Pi_i(z, M) = p_i(1-(1-q_i)^M)$. Then we have the following two lemmas:

Lemma A.1. *For any demand level M and total attractiveness $z_1 > z_2 > z_i^*$, where z_i^* is the attractiveness level when we only display the i-th seller under demand M,*

we get

$$\frac{|\Pi_i(z_1, M) - \Pi_i(z_2, M)|}{\Pi_i(z_2, M)} \leq \frac{1 + q_i(z_2, M)}{\beta p_i(z_2, M)} \frac{z_1 - z_2}{1 + z_2} \leq \frac{2(z_1 - z_2)}{1 + z_2}. \tag{A.14}$$

Here, $q_i(z_2, M)$ and $p_i(z_2, M)$ are the corresponding probability and price under z_2 and M.

Lemma A.2. *For any demand level $M_1 > M_2$ and total attractiveness level $z > z_i^*$, where z_i^* is the attractiveness level when we only display the i-th seller under demand M_1, we get*

$$\frac{|\Pi_i(z, M_1) - \Pi_i(z, M_2)|}{\Pi_i(z, M_2)} \leq \left(\frac{1}{\beta p_i(z, M_2)} + q_i(z, M_2) \right) \frac{M_2 - M_1}{M_2} \leq \frac{2(M_2 - M_1)}{M_2}. \tag{A.15}$$

Here, $p_i(z, M_2)$ and $q(z, M_2)$ are the corresponding price and probability under z and M_2.

Proof for Proposition 2.3: We will use these two lemmas to prove the statement in Proposition 2.3. Suppose seller i is in a partition of attractiveness z_1 and demand M_1 and he/she hope to switch to another partition with attractiveness z_2 and demand M_2. We further assume after the switching, the attractiveness of the new partition becomes z_2^*. Obviously, we get $z_2^* > z_2$. So the envy level can be written as

$$EN_i = \frac{\Pi_i(z_2^*, M_2) - \Pi_i(z_1, M_1)}{\Pi_i(z_1, M_1)}.$$

If $z_1 < z_2$ or $M_1 > M_2$, then we can directly apply Lemma A.2 or A.1 with replacing z_2 to be z_1 or M_2 to be M_1. Hence, we only need to prove the case where $z_1 \geq z_2$ and $M_1 \leq M_2$. Denote $\Pi_i^1 = \Pi_i(z_1, M_1), \Pi_i^2 = \Pi_i(z_2^*, M_1), \Pi_i^3 = \Pi_i(z_2^*, M_2)$, then from Lemma A.1:

$$\frac{\Pi_i^2 - \Pi_i^1}{\Pi_i^2} \leq 2\frac{z_1 - z_2^*}{1 + z_2^*} \leq 2\frac{z_1 - z_2}{z_2} \leq 2(\frac{1}{\alpha} - 1), \tag{A.16}$$

where the last inequality follows from the definition of α-fair policy. Similarly, by

Lemma A.2, we get

$$\frac{\Pi_i^3 - \Pi_i^2}{\Pi_i^2} \leq 2\frac{M_2 - M_1}{M_1} \leq 2(\frac{1}{\delta} - 1), \tag{A.17}$$

where the last inequality follows from the definition of δ-fair policy. Furthermore, if $2/3 < \alpha$, then $2(\frac{1}{\alpha} - 1) < 1$, so by using (A.16), we could obtain

$$\frac{\Pi_i^2}{\Pi_i^1} \leq \frac{\alpha}{3\alpha - 2}.$$

Finally, by adding (A.16) and (A.17), and noticing the inequality above, we would have

$$\frac{\Pi_i^3 - \Pi_i^1}{\Pi_i^1} \leq \frac{2\alpha}{3\alpha - 2}\left(\frac{1}{\alpha} + \frac{1}{\delta} - 2\right).$$

Proof for Lemma A.1: We will use mean value theorem to estimate the difference in revenue when the attractiveness z moves from z_1 to z_2. In this section, we omit the same demand level M for simplicity. We first consider $\frac{\partial p_i}{\partial z}$. From the definition of $q_i = \frac{e^{a_i - \beta p_i}}{1+z}$, we get $\frac{\partial q_i}{\partial z} = q_i(-\beta\frac{\partial p_i}{\partial z} - \frac{1}{1+z})$. By differentiating both side of the FOC equation and plugging in this relationship, we get

$$\frac{\partial p_i}{\partial z} = \frac{\beta p_i(1 - q_i) - (1 + M\beta p_i q_i)}{\beta(1+z)\left((1 - \beta p_i)(1 - q_i) + 1 + M\beta p_i q_i\right)}. \tag{A.18}$$

Then, because $\Pi_i(z) = p_i - \frac{p_i}{M\beta p_i q_i + 1}$, we further achieve

$$\frac{\partial \Pi_i(z)}{\partial z} = -\frac{\Pi_i}{\beta p_i(1+z)}\left(1 + \frac{q_i}{2 + M\beta p_i q_i - \beta p_i + \beta p_i q_i - q_i}\right). \tag{A.19}$$

Note that by applying the Bernoulli inequality to the FOC, we can get

$$1 = (1 + M\beta p_i q_i)(1 - q_i)^M > (1 + M\beta p_i q_i)(1 - Mq_i) \Longrightarrow 1 + M\beta p_i q_i > \beta p_i. \tag{A.20}$$

And by taking logarithm of the FOC and using the fact that $\ln(1+x) \leq x$ for $x > -1$, we will get a lower bound for p_i:

$$0 = \ln(1 + M\beta p_i q_i) + M\ln(1 - q_i) \leq M\beta p_i q_i - Mq_i \Longrightarrow \beta p_i \geq 1. \tag{A.21}$$

Now that we have (A.20) and (A.21), we can estimate the denominator in equation (A.19) that $2 + M\beta p_i q_i - \beta p_i + \beta p_i q_i - q_i > 1$. Hence,

$$-\frac{\Pi_i}{\beta p_i(1+z)}(1+q_i) < \frac{\partial \Pi_i(z)}{\partial z} < -\frac{\Pi_i}{\beta p_i(1+z)}. \tag{A.22}$$

From (A.20) and (A.18), we can conclude that $\frac{\partial p_i}{\partial z} < 0$, which means as the total attractiveness increases, the price p_i will decrease in return. And for probability q_i, we can calculate

$$\frac{\partial q_i}{\partial z} = -\frac{q_i}{1+z} * \frac{1-q_i}{((1-\beta p_i)(1-q_i)+1+M\beta p_i q_i)} < 0. \tag{A.23}$$

The term $\frac{\Pi_i}{p_i} = 1-(1-q_i)^M$ is therefore decreasing as z increases as well. Hence this is also the case for $\frac{\Pi_i}{p_i(1+z)}$ and $\frac{\Pi_i(1+q_i)}{p_i(1+z)}$. Then by mean value theorem, for any $z_1 > z_2$, there exists a $z_2 < z^* < z_1$ such that

$$\Pi_i(z_1)-\Pi_i(z_2) = \frac{\partial \Pi_i(z^*)}{\partial z^*}(z_1-z_2) > -\frac{\Pi(z^*)(1+q_i(z^*))}{\beta p_i(z^*)(1+z^*)}(z_1-z_2) > -\frac{\Pi_i(z_2)(1+q_i(z_2))}{\beta p_i(z_2)(1+z_2))}(z_1-z_2).$$

which further gives (note that $z_1 > z_2$ means $\Pi_i(z_1) < \Pi_i(z_2)$)

$$\frac{|\Pi_i(z_1)-\Pi_i(z_2)|}{\Pi_i(z_2)} < \frac{1+q_i(z_2)}{\beta p_i(z_2)}\frac{z_1-z_2}{1+z_2} < \frac{2(z_1-z_2)}{1+z_2}, \tag{A.24}$$

where the last inequality follows from $\beta p_i > 1$ and $q_i < 1$.

Proof for Lemma A.2: For the demand M, we use similar approach as in lemma A.1. First, by the implicit function theorem, we are able to calculate $\frac{\partial p}{\partial M}$ as:

$$\frac{\partial p_i}{\partial M} = \frac{(\beta p_i q_i + \ln(1-q_i)(1+M\beta p_i q_i))(1-q_i)}{M\beta q_i((\beta p_i - 1)(1-q_i)-(1+M\beta p_i q_i))}. \tag{A.25}$$

By using (A.20), we can easily see that $\frac{\partial p_i}{\partial M} > 0$. Meanwhile, we can also derive $\frac{\partial \Pi_i}{\partial M}$

by implicit function theorem:

$$\frac{\partial \Pi_i}{\partial M} = -\frac{(\beta p_i q_i + (1+M\beta p_i q_i)\ln(1-q_i))(1-q_i)(M\beta p_i q_i + 2 - \beta p_i)}{(M\beta p_i q_i+1)^2[(1-\beta p_i)(1-q_i)+1+M\beta p_i q_i]} + \frac{p_i^2 q_i}{(1+M\beta p_i q_i)^2} \tag{A.26}$$

$$= p_i \frac{-(1-q_i)\ln(1-q_i)(M\beta p_i q_i + 2 - \beta p_i) + \beta p_i q_i^2}{(M\beta p_i q_i+1)[(1-\beta p_i)(1-q_i)+1+M\beta p_i q_i]}. \tag{A.27}$$

Because we have $M\beta p_i q_i + 1 > \beta p_i$ and $\ln(1-q_i) < -q_i$, then the first term in (A.26) is positive, which gives us $\frac{\partial \Pi_i}{\partial M} > 0$.

For the fact that $0 < q < 1$, then by simple calculus, we can show that $|(1-q)\ln(1-q)| \leq q$, so we are able to estimate $|\frac{\partial \Pi_i}{\partial M}|$ as

$$\begin{aligned}
|\frac{\partial \Pi_i}{\partial M}| =& p_i \frac{|(1-q_i)\ln(1-q_i)(M\beta p_i q_i + 2 - \beta p_i)| + \beta p_i q_i^2}{(M\beta p_i q_i+1)[(1-\beta p_i)(1-q_i)+1+M\beta p_i q_i]} \\
<& p_i \frac{q_i(M\beta p_i q_i + 2 - \beta p_i) + \beta p_i q_i^2}{(M\beta p_i q_i+1)[(1-\beta p_i)(1-q_i)+1+M\beta p_i q_i]} \\
<& \frac{p_i}{M\beta p_i q_i + 1}[q_i + \frac{q_i(q_i(1-\beta p_i))}{(1-\beta p_i)(1-q_i)+1+M\beta p_i q_i} + \beta p_i q_i^2] \\
<& \frac{p_i}{M\beta p_i q_i + 1}[q_i + \beta p_i q_i^2] = \frac{\Pi_i}{M}(\frac{1}{\beta p_i} + q_i)
\end{aligned} \tag{A.28}$$

We next prove the term $\frac{p_i}{M\beta p_i q_i + 1}$ is decreasing with respect to M. If we have the mononicity condition, note that $p_i q_i^2$ is also decreasing with respect to M by the fact that $\beta p_i > 1$ in (A.21), then the term $\frac{\Pi_i}{M}(\frac{1}{\beta p_i} + q_i)$ is decreasing with respect to M. By (A.25) and (A.26), we can calculate the derivative as

$$\frac{\partial(\frac{p_i}{M\beta p_i q_i+1})}{\partial M} = \frac{(1+M\beta p_i^2 q_i)\frac{\partial p_i}{\partial M} - \beta^2 p_i^2 q_i}{\beta(M\beta p_i q_i+1)^2}. \tag{A.29}$$

Hence $\frac{\partial(\frac{p_i}{M\beta p_i q_i+1})}{\partial M} < 0 \iff \frac{\partial p_i}{\partial M} < \frac{\beta^2 p_i^2 q_i}{1+M\beta^2 p_i^2 q_i}$. By plugging in (A.25), this condition is further equivalent to

$$\beta p_i q_i(1-q_i) + M\beta^2 p_i^2 q_i^2 + \ln(1-q_i)(1+M\beta^2 p_i^2 q_i)(1-q_i) > 0.$$

By simple calculus, we can prove $x\ln(x) > x-1$ for all $1 > x > 0$. Hence, $\ln(1-q_i)(1-$

$q_i) > -q_i$. As a result, we only need to prove $\beta p_i q_i + \ln(1-q_i) > 0$. This can be derived from the FOC equation. Recall that the FOC is $\ln(1 + M\beta p_i q_i) + M\ln(1 - q_i) = 0$, so $\ln(1 - q_i) = -\frac{\ln(1+M\beta p_i q_i)}{M} > -\beta p_i q_i$ where we use the fact that $\ln(1 + x) < x$ for $x > -1$.

Hence, for any $M_1 > M_2$, by mean value theorem, there exists $M_1 > M^* > M_2$ such that

$$\Pi_i(M_1) - \Pi_i(M_2) = |\frac{\partial \Pi_i(M^*)}{\partial M}|(M_1 - M_2) \leq |\frac{\partial \Pi_i(M_1)}{\partial M}|(M_1 - M_2) \leq \Pi_i(M_1)\left(\frac{1}{\beta p_i(M_1)} + q_i(M_1)\right)\frac{M_1 -$$

where we used the monotonicity condition for the partial derivative $\frac{\partial \Pi}{\partial M}$. Finally, by rearranging terms, we have

$$\frac{\Pi_i(M_1) - \Pi_i(M_0)}{\Pi_i(M_i)} \leq \left(\frac{1}{\beta p_i(M_1)} + q_i(M_1)\right)\frac{M_1 - M_2}{M_1} \leq 2\frac{M_1 - M_2}{M_1}.$$

Here, the last inequality follows from $\beta p_i > 1$ and $q_i < 1$.

A.1.8 Proof for Proposition 2.4

We first prove that under the equilibrium, we have $p_1 \geq p_2 \geq \cdots \geq p_N > 1$, $q_1 \geq q_2 \geq \cdots \geq q_N$ and $e^{a_1 - \beta p_1} \geq e^{a_2 - \beta p_2} \geq \cdots \geq e^{a_N - \beta p_N}$, i.e., products with higher quality should charge higher prices and have larger attractiveness. From the first order condition, we have $\frac{1}{\beta p_i} = 1 - q_i$, which indicates $\beta p_i > 1$ immediately. Then for any $a_i \geq a_j$

$$\frac{1 - 1/(\beta p_i)}{1 - 1/(\beta p_j)} = \frac{q_i}{q_j} = e^{(a_i - \beta p_i) - (a_j - \beta p_j)} \geq e^{\beta(p_j - p_i)}$$

if $p_i < p_j$, then the left side $\frac{1-1/(\beta p_i)}{1-1/(\beta p_j)} < 1$ and the right side $e^{\beta(p_j - p_i)} > 1$, which is a contradiction. Hence we must have $p_i \geq p_j$. From the FOC, this indicates $q_i \geq q_j$ which further shows $e^{a_i - \beta p_i} \geq e^{a_j - \beta p_j}$.

Then for any $i = 1, 2, \ldots, N$, we have

$$\frac{1}{\beta p_i} = 1 - \frac{e^{a_i - \beta p_i}}{1 + \sum_k e^{a_k - \beta p_k}} \geq 1 - \frac{e^{a_i - \beta p_i}}{1 + N e^{a_N - \beta p_N}}. \tag{A.30}$$

Denote p_N^* to be the solution of equation

$$\frac{1}{\beta p_N^*} = 1 - \frac{e^{a_N - \beta p_N^*}}{1 + N e^{a_N - \beta p_N^*}},$$

then we have $p_N < p_N^*$ by (A.30). We can also easily see that $\lim_{N\to\infty} \beta p_N^* = 1$

For the seller $i = 1$, we further have

$$\frac{1}{\beta p_1} \geq 1 - \frac{e^{a_1 - \beta p_1}}{1 + N e^{a_N - \beta p_N}} \geq 1 - \frac{e^{a_1 - \beta p_1}}{1 + N e^{a_N - \beta p_N^*}}.$$

Let p_1^* be the solution that satisfies

$$\frac{1}{\beta p_1^*} = 1 - \frac{e^{a_1 - \beta p_1^*}}{1 + N e^{a_N - \beta p_N^*}}, \tag{A.31}$$

then we have $p_N \leq p_{N-1} \leq p_{N-2} \leq \cdots \leq p_1 \leq p_1^*$. As a result, $\Pi(\mathcal{S}) = \sum_{i=1}^N p_i q_i = \sum_{i=1}^N p_i - N/\beta \leq N(p_1^* - 1/\beta)$. And note that $\lim_{N\to\infty} p_1^* = \lim_{N\to\infty} p_N^* = 1/\beta$, this enable us to derive an upper bound on the revenue $\Pi(\mathcal{S})$

$$\Pi(\mathcal{S}) < N(p_1^* - 1/\beta) = p_1^* \frac{N e^{a_1 - \beta p_1^*}}{1 + N e^{a_N - \beta p_N^*}} \to e^{a_1 - a_N}/\beta, \quad (N \to \infty).$$

For a single display with product a_1, the revenue is $\Pi(\{a_1\}) = p'q' = p' - 1/\beta$, where p' satisfies

$$\frac{1}{\beta p'} = 1 - \frac{e^{a_1 - \beta p'}}{1 + e^{a_1 - \beta p'}} \iff p' - \frac{1}{\beta} = \frac{e^{a_1 - p'}}{\beta}.$$

So $\Pi(a_1) > \lim_{N\to\infty} \Pi(\mathcal{S}) \iff a_N > p' \iff a_N > (1 + e^{a_1 - a_N})/\beta$, which is satisfied by the condition $u - l < \ln(\beta u - 1)$.

A.2 Simplification of the Main Formulation

The idea behind the alternative formulations is to reduce the dimension of the problem. To achieve this, we remove the subscript k from formulation (2.7) and characterize each partition through a unique (j, v) dyad. Specifically, $z_{j,v}$ specifies whether the

partition with total attractiveness $\mathcal{Z}_j$ is assigned with market share P_v. In addition, we denote by $x_{i,j,v}$ a binary decision variable indicating whether to allocate $\mathcal{Z}_j$ and market share P_v to seller i, and there are in total K pairs of (j, v) combinations. The value $x_{i,j,v}$ will then shape the first constraint and the total revenue through input tables $E_{i,j,v}$ and $\Pi_{i,j,v}$. The specific formulation is as follows:

$$
\begin{aligned}
\textbf{(LB)} \qquad & \max_{x,z} \sum_{k} x_{i,j,v}\, \Pi_{i,j,v} && && \text{(A.32)} \\
& s.t. \sum_{i} x_{i,j,v} E_{i,j,v} = z_{j,v} \mathcal{Z}_j && \forall j, v \\
& \sum_{j,v} x_{i,j,v} = 1, && \forall i \\
& \sum_{j,v} z_{j,v} = K, \\
& \sum_{j,v} z_{j,v} P_v = 1, \\
& x_{i,j,v} \text{ binary}, \quad z_{j,v} \text{ binary}
\end{aligned}
$$

In fact, formulation (2.7) (denoted Main) is a relaxation of formulation (A.32) (denoted LB) and the optimal total revenue of LB serves as a lower bound of that for Main. To see this, note that any feasible solution of LB is still feasible under Main. However, a feasible solution in Main in which two partitions receive the same (j, v) pair is infeasible in LB, as this will make the corresponding $z_{j,v} = 2$, thereby violating the binary constraint.

Using a similar idea, we construct another formulation (UB) whose objective is the upper bound of the main formulation. Specifically, the formulation of UB is identical to that of LB, except that we require $z_{j,v}$ to be nonnegative integers instead of binary variables. Again, using the same argument from the previous paragraph, it is clear that any feasible solution in the main formulation is also feasible for UB, but integer solutions such as $z_{j,v} = 2$ are feasible only in UB.

These two formulations, LB and UB, serve to provide the lower and upper bounds to the main optimization problem. Such approximations are particularly useful when

solving the main problem turns out to be computationally challenging, which is the case when the number of listings becomes large and the discretization gap becomes sufficiently small. In Figure A-1, we show the revenue and running time of the three formulations, together with a fourth option, namely, the LP relaxation of the main formulation. The instance is the same as in the previous numerical analysis. It is observed that the upper and lower bound and the LP relaxation yield almost the same outcome as the main formulation. However as shown in panel (a), we need finer discretization (smaller step sizes) of $\mathcal{Z}$ for the formulation to have a good approximation. However, panel (b) shows that when the step size is very small, the lower and upper bound formulations will have significantly shorter computational time.

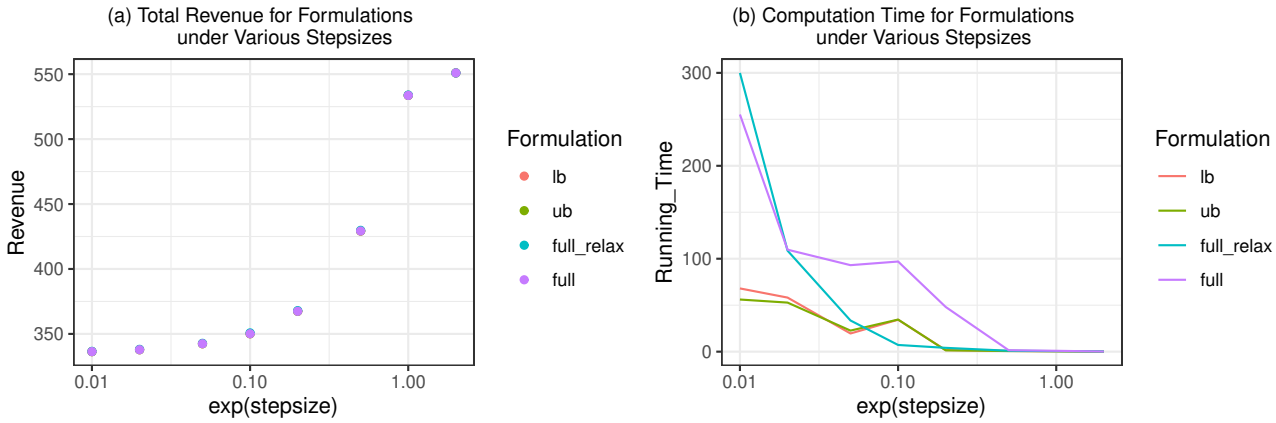

Figure A-1: Revenue and Running Time under Partitioned Display

Appendix B

Proofs and Supplemental Material for Chapter 3

B.1 Background Information on Airbnb

The labelling task is undertaken by four subjects. Similar to Zhang et al. (2017), we prepare a detailed scoring instruction in which we give examples of images that have score from 1 to 7 where higher score indicates a higher image quality. Specifically, image score is determined by how attractive the image is. For example, a score 1 is given to the image in panel (a) of Figure B-1 as the photo is underexposed and doesn't accurately represent the room features, while a score 7 is assigned to panel (b) of the Figure B-1 since the exposure, lighting and hue of the image are all well controlled and it clearly represents the kitchen.

There can be multiple reasons that lead to inaccurate room type classification. First, it is possible for an image to capture a view of multiple room types. For example, the picture in panel (a) of the Figure B-2 shows both bedroom and living room in the same picture, which adds uncertainty into both the labeling and classification processes. In this case, we ask our subjects to label the main room type of the images, and when the labels are not unanimous, we label the room type that the majority chooses (with a random draw in the case of a tie). Second, the images do not belong to any of the five room types (such as fitness rooms). When labeling the images in the

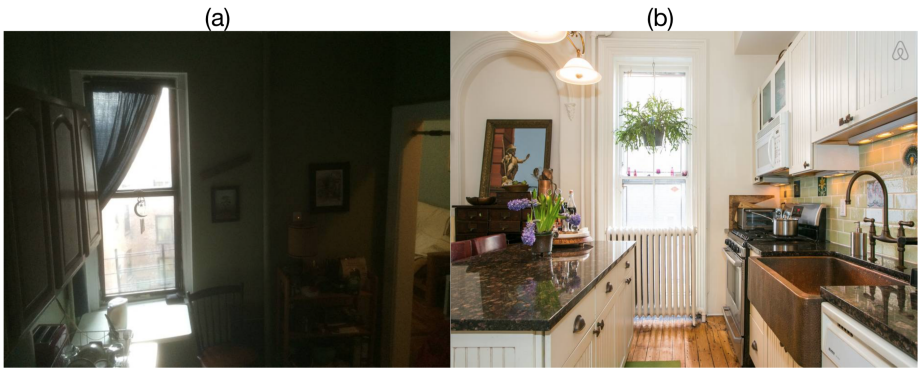

Figure B-1: Quality variation in photos

training set for the CNN model, subjects are asked to mark these images as NA, and they are excluded from the final training data. While it is possible that images of the NA type could appear in the test set and be classified as one of the five room types, we would like to note that the proportion of NA type images in the first five images on the display web page is very small - they only account for 1.05% of the training set. Including the NA type will generate more noise in the classification process, as NA may include any random room or subject, while the insufficient number of images makes it almost impossible to train the neural network to recognize all the patterns. Third, objects in different rooms may look identical to each other. For example, the image in panel (b) of Figure B-2 is mis-classified as bedroom as the sofa looks similar to a bed.

Overall, our current ResNet50 model achieves an accuracy of 84%. To investigate the impact of photo type mis-specification on estimation outcomes, in the estimation stage, we perturb the actual room types into the mis-classified types using the probability transition matrix presented in Figure 3-2. The detailed setup and estimation results are summarized in Appendix B.2.

B.2 Robustness Check for Estimation Results

In Table 3.7 the estimation results assume Gumbel error term distribution in the linear utility model and photo layout defined as in 3.2.3. In this appendix section we present

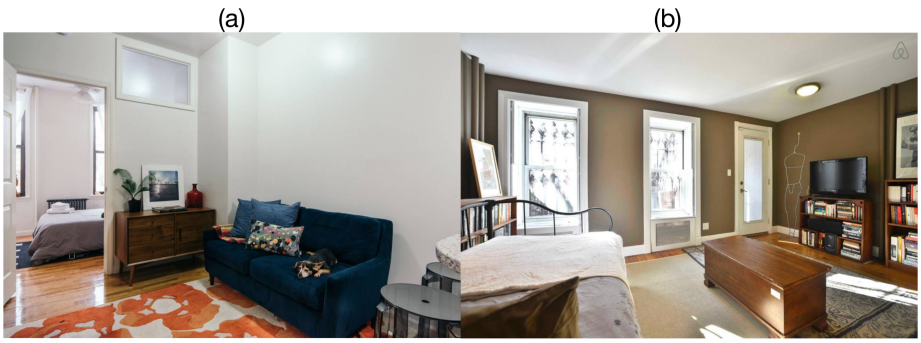

Figure B-2: Misspecification of Bedrooms

several alternative specifications to confirm the robustness of estimation results, over the same regions and time. To start with, we examine (1) the estimation results using GLM instead of NLS on the same specification, (2) the estimation results using photo scores constructed from objective measures, and (3) the potential non-linearity effect from photo layout. The estimation results are presented in Tabel B.1.

Estimation Results using GLM We estimate the same specification using Equation 3.9 and present the results in column (3) of Table B.1. Compared to the estimates obtained through NLS, the relative order of impact from each photo type remains the same, i.e., the coefficient for bedroom photo still has the largest magnitude among all the room types. However, when the error term is mis-specified, the GLM estimation results on synthetic data in Table 3.6 demonstrate larger deviations from the true vale. As a result, we use the NLS estimation results in the main text, despite the significant reduction on computation time of the GLM approach.

Fitted Photo Quality In addition to running GLM using the labeled photo quality, we rerun our estimation using fitted photo quality, which is comprised by six objective photo quality measures, namely Hue, Saturation, Brightness, Contrast, Clarity and Resolution (the construction of the fitted photo quality variable is provided in the Online Companion). The estimation results are shown in column (1) of Table B.1. When using fitted photo quality, the impact from cover photos become larger specification (3). However, in both cases, bedroom covers lead to the highest increase in apartment attractiveness.

The Nonlinearity Effect from Photo Layout We further add squared terms for

Table B.1: PCM (GLM) Estimation Results under Different Photo Quality Specifications

	Fitted Photo Quality		Labled Photo Quality	
	(1)	(2)	(3)	(4)
Price	−0.342*** (0.001)	−0.337*** (0.001)	−0.363*** (0.001)	−0.358*** (0.001)
Overall_Rating	0.319*** (0.003)	0.322*** (0.003)	0.287*** (0.003)	0.294*** (0.003)
Number_of_Reviews	0.007*** (0.00002)	0.008*** (0.00002)	0.007*** (0.00002)	0.008*** (0.00003)
Airbnb_Superhost	0.196*** (0.003)	0.199*** (0.003)	0.188*** (0.003)	0.196*** (0.003)
Response_Rate	0.008*** (0.0001)	0.007*** (0.0001)	0.008*** (0.0001)	0.008*** (0.0001)
Num_of_photo	0.002*** (0.0001)	0.003*** (0.0001)	0.001*** (0.0001)	0.002*** (0.0001)
Num_of_Bedrooms	0.176*** (0.002)	0.173*** (0.002)	0.166*** (0.002)	0.165*** (0.002)
Num_of_Bathrooms	0.234*** (0.005)	0.201*** (0.005)	0.239*** (0.005)	0.197*** (0.005)
Photo: Cover (a_m^c)				
Bedroom	0.193*** (0.003)	0.178*** (0.003)	0.119*** (0.002)	0.112*** (0.002)
Living Room	0.163*** (0.003)	0.149*** (0.003)	0.096*** (0.002)	0.093*** (0.002)
Outside	0.188*** (0.003)	0.169*** (0.003)	0.114*** (0.003)	0.105*** (0.003)
Kitchen	0.181*** (0.003)	0.167*** (0.003)	0.087*** (0.003)	0.082*** (0.003)
Bedroom^2		0.003*** (0.0002)		0.001*** (0.0002)
Living Room^2		0.004*** (0.0002)		0.004*** (0.0002)
Outside^2		0.007*** (0.0003)		0.004*** (0.0004)
Kitchen^2		0.006*** (0.0002)		0.007*** (0.0003)
Photo: Non-Cover (a_m^{nc})				
Bedroom	0.021*** (0.001)	0.018*** (0.001)	0.027*** (0.001)	0.023*** (0.001)
Living Room	0.013*** (0.001)	0.012*** (0.001)	0.022*** (0.001)	0.021*** (0.001)
Outside	0.009*** (0.001)	0.010*** (0.001)	0.011*** (0.001)	0.009*** (0.001)
Toilet	0.006*** (0.001)	0.002 (0.001)	0.016*** (0.001)	0.010*** (0.001)
Kitchen	0.021*** (0.001)	0.020*** (0.001)	0.025*** (0.001)	0.023*** (0.001)
Bedroom^2		0.0004*** (0.00002)		0.0005*** (0.00002)
Living Room^2		0.0003*** (0.00002)		0.0003*** (0.00002)
Outside^2		0.002*** (0.0001)		0.001*** (0.0001)
Toilet^2		0.002*** (0.0001)		0.003*** (0.0001)
Kitchen^2		0.001*** (0.00004)		0.001*** (0.00004)
Photo: Duplicate (δ_m)				
Bedroom	−0.004*** (0.001)	−0.003*** (0.001)	−0.007*** (0.001)	−0.004*** (0.001)
Living Room	0.009*** (0.001)	0.011*** (0.001)	0.004*** (0.001)	0.004*** (0.001)
Observations	1,566,415	1,566,415	1,395,827	1,395,827
Note:				*p<0.1; **p<0.05; ***p<0.01

the scores of cover and non-cover photos to capture potential non-linear effect. As shown in Column (2) and (4) of Table B.1, the magnitudes for the second order terms are very small, while the rest of the parameters are similar in magnitudes to other specifications.

Robustness Check for Misspecified Consideration Set A new apartment may pop up at any time, or an existing listing may switch its status between "B" (Blocked) and "A" (Available), so that the consideration set is not strictly shrinking over time.

Table B.2: PCM (GLM) Estimation Results under Misspecified Listing Status and Image Types through Simulated Synthetic Data

	Dependent variable:			
	Real Value	Unadjusted	Adjusted	Misspecified Image Types
Price	−0.402	−0.404***	−0.404***	-0.401***
		(0.0001)	(0.0001)	(0.0037)
Overall.Rating	0.307	0.306***	0.305***	0.305***
		(0.0002)	(0.0002)	(0.0059)
Number.of.Reviews	0.007	0.007***	0.007***	0.007***
		(0.000001)	(0.000001)	(0.000005)
Airbnb.Superhost	0.186	0.190***	0.190***	0.185***
		(0.0002)	(0.0002)	(0.0063)
Response.Rate	0.008	0.008***	0.008***	0.008***
		(0.000004)	(0.000004)	(0.000178)
Num_of_Photos	0.001	0.001***	0.001***	0.001***
		(0.00001)	(0.00001)	(0.00033)
Num_of_Bedrooms	0.228	0.229***	0.229***	0.228***
		(0.0001)	(0.0001)	(0.0046)
Num_of_Bathrooms	0.106	0.114***	0.115***	0.109***
		(0.0002)	(0.0002)	(0.0110)
Cover				
Bedroom	0.110	0.112***	0.111***	0.093***
		(0.0001)	(0.0001)	(0.0038)
Living Room	0.080	0.082***	0.081***	0.074***
		(0.0001)	(0.0001)	(0.0035)
Outside	0.080	0.082***	0.081***	0.070***
		(0.0001)	(0.0001)	(0.0038)
Kitchen	0.090	0.090***	0.090***	0.079***
		(0.0001)	(0.0001)	(0.0036)
Non-Cover				
Bedroom	0.030	0.030***	0.031***	0.030***
		(0.00003)	(0.00003)	(0.00134)
Living Room	0.020	0.020***	0.020***	0.023***
		(0.00003)	(0.00003)	(0.00148)
Outside	0.010	0.011***	0.011***	0.012***
		(0.00004)	(0.00004)	(0.00161)
Toilet	0.020	0.020***	0.020***	0.022***
		(0.00004)	(0.00004)	(0.00159)
Kitchen	0.020	0.020***	0.020***	0.022***
		(0.00003)	(0.00003)	(0.00146)
Duplicate				
Bedroom	−0.010	−0.010***	−0.011***	-0.007***
		(0.00003)	(0.00003)	(0.00156)
Living Room	0.006	0.006***	0.006***	0.004***
		(0.00003)	(0.00003)	(0.00148)
Observations		1,846,081	1,846,081	
Note:			*p<0.1; **p<0.05; ***p<0.01	

In our main analysis in Table 3.7, we address this issue by focusing the demand within one month before the check-in date, as listing owners are less likely to switch the status of their listing as the check-in date approaches. Without observing the status trajectory, we assume that these trajectories are consistent with the status documented in our dataset. Specifically, on any given day, if a listing ends up not

being booked (available), then we assume it has been available for the past month. Similarly, if a listing is blocked in the final status, we assume it has been blocked for the past month as well. To test the validity of our assumption, we acquired a new dataset that documents the trajectory of the status of each listing from February 1st to May 1st. Specifically, we observe listing i's availability trajectory on a specific check-in date, let us say, March 1st, from February 1st all the way to March 1st. In this way, we are able to check whether the owner of property i changes the status of the listing before the check-in date or not.

Importantly, we examine the percentage of miscounts caused by irregular variation in the choice set, which could result from the following three cases: 1) a new apartment is listed on the website halfway, but is counted as available throughout the month in our previous analysis, 2) an existing listing that switches between "B" and "A" but is counted as "A" in the previous estimation process, and 3) an existing listing that switches between "B" and "A" but is counted as "B" in the previous estimation process.

Cases 1 and 2 are similar, as they both refer to situations where a listing was blocked for some days but was considered available when we count the number of comparisons. In these cases, as the listing did not participate in all the comparisons, the number of comparisons is overestimated. Case 3 is the opposite of Cases 1 and 2 in that the listing was made available for some days but was not accounted for in our previous analysis. In this case, we underestimated the number of comparisons. For each listing i, the percentage of miscounts is calculated by dividing the total number of violations for listing i (i.e., number of days/incidences that satisfy the three cases mentioned above) across the timespan of our dataset by the total number of statuses for listing i tracked in our dataset. We find that scenarios 1 and 2 combined account for 0.83% of the total number of cases, and scenario 3 accounts for 0.41% of the total number of cases. In total, all three cases comprise 1.24% of all the observations. We also investigate the impact of miscounted cases on our estimation results through synthetic data (as our availability data span only three months, we are not able to tune our previous estimation using real data that includes transactions over the

entire year). For each listing, we record the number of miscounts and correct the bias caused by the abovementioned cases. We present the estimation results in Table B.2. The estimation results are almost identical under the adjusted and unadjusted specifications, suggesting that when the percentage of miscounts is low, assuming that the offer set is shrinking has a very limited impact on the estimation results.

Robustness Check for Misspecified Image Type The output from CNN is subject to misclassifications (currently the state-of-the-art classification has approximately 88% accuracy[1]). In our case, Figure B-2 provides a breakdown of the by-room-type accuracy from our ResNet50 model, which can potentially bias the estimation results. As it is not possible to manually recover the true type for every one of the 222,144 photos, we test the impact of photo type misspecification on the estimation outcomes using synthetic data. To best approximate the actual data-generating process, we sample 1,000 listings from our data. We then generate the room type of each image according to the approximated empirical distribution (B=30%, L=30%, O=10%, K=20%, T=10%). Importantly, the underlying coefficient of each feature is assumed to align with our GLM estimation results. In this way, the mean attractiveness of each listing is computed by plugging in the estimation results in Table B.2 into Equation 3.5. With the assumption that the error term in Equation 3.5 follows a Gumbel distribution and that the outside options are normalized to 0, we simulate the transaction data for the 1,000 listings in a timespan of one year.

In the estimation stage, we perturb the actual room types into the misclassified types using the probability transition matrix presented in Figure 2. The estimation results are summarized in Table B.2. We observe that misclassification in our simulated data generates downward bias only for the coefficients of cover effects. Nevertheless, the order of impact for different room types remains the same, i.e., all else remaining equal, a bedroom photo is still the best candidate for the cover. This is possibly because the percentage of misspecified photos does not vary too much across room types. Other than for the cover image, photo type misclassification appears to have a very limited impact on the coefficients for features. As a result, failing to account

[1]https://paperswithcode.com/sota/image-classification-on-imagenet

for photo type misclassification will understate the revenue gain when each listing switches to the optimal photo layout in the counterfactual analysis, as the increase in the overall attractiveness from the optimal photo layout should actually be higher under the correctly specified scenario.

B.3 Numerical Experiment using Synthetic Data

We perform estimation through SMNL, PCM (NLS) and PCM (GLM) under different true values of the parameter of interest, β_0. Similar to the data generating process we use for Table 3.6, we assume that there are 500 potential customers. Upon each customer's arrival, an i.i.d. error term is realized for each listing, and the listing with the highest realized utility is booked. If every listing's realized utility is lower than that of the outside option (whose mean utility is normalized to zero), then that demand is lost. Throughout the numerical experiment, we observe the decision made by each customer, i.e., if the customer chooses a specific listing or the outside option. We first perform estimation without observing the no-purchase data. Then, to investigate how the absence of no-purchase data affects our estimation results, we repeat the analysis by incorporating no-purchase data into our comparisons. We use "+O" to signify such specifications in Tables B.1 and B.2. Under such specifications, the PCM model compares all pairs of options including the outside option. Unlike listings on Airbnb that have unit availability, the outside option has almost unlimited inventory and therefore can be selected multiple times on a specific day. Nevertheless, the counting principle of PCM remains the same. The number of times that the outside option o is preferred to option i on day d, $\mathbb{1}_{\{o \succ i\}d}$, is counted by the number of times on day d that a customer chooses o when i is also available. For instance, on a specific day, if 10 customers ultimately book the outside option until listing i is booked, then PCM will count 10 times for the term $\mathbb{1}_{\{o \succ i\}d}$ and one time for the term $\mathbb{1}_{\{i \succ o\}d}$. In this way, we can construct the comparison matrix similarly as before, and the estimation follows the standard PCM (either NLS or GLM) procedures. Tables B.1 and B.2 report the estimation results under $\beta_0 = 0.5$ and $\beta_0 = 2$, respectively.

We observe that PCM is able to uncover the ground-truth parameter regardless of whether outside option observations are included.

Table B.1: Extended Estimation Results Using Synthetic Data ($\beta_0 = 0.5$, 95% CI)

Error Structure	Gumbel			Mixed-Normal		
Choice Model	Mean	St. Dev.	RMSE	Mean	St. Dev.	RMSE
SMNL	0.499	0.012	0.012	0.610	0.032	0.115
PCM (NLS)	0.501	0.014	0.014	0.535	0.018	0.040
PCM (GLM)	0.501	0.016	0.016	0.584	0.037	0.092
PCM+O (NLS)	0.500	0.012	0.012	0.530	0.016	0.035
PCM+O (GLM)	0.498	0.015	0.015	0.572	0.026	0.077

Table B.2: Extended Estimation Results Using Synthetic Data ($\beta_0 = 2$, 95% CI)

Error Structure	Gumbel			Mixed-Normal		
Choice Model	Mean	St. Dev.	RMSE	Mean	St. Dev.	RMSE
SMNL	1.989	0.027	0.028	2.357	0.073	0.364
PCM (NLS)	1.988	0.036	0.038	2.147	0.048	0.154
PCM (GLM)	1.981	0.050	0.054	2.305	0.125	0.329
PCM+O (NLS)	2.002	0.035	0.035	2.131	0.072	0.149
PCM+O (GLM)	1.979	0.046	0.050	1.935	0.162	0.173

B.4 Proof for Proposition 3.1 - 3.3

Proof for Proposition 3.1 - 3.2. To prove consistency nonlinear regression model, we need to verify four conditions of the consistency Theorem 4.3.1 in Amemiya (1985). That is, for any nonlinear regression that takes form of $y_t = f_t(\boldsymbol{\beta}) + \epsilon_t$, we have:

(A) $\partial f_t / \partial \boldsymbol{\beta}$ exists and is continuous on N.

(B) $f_t(\boldsymbol{\beta})$ is continuous in $\boldsymbol{\beta} \in N$ uniformly in t.

(C) $\dfrac{1}{T}\sum_{t=1}^{T} f_t(\boldsymbol{\beta_1}) f_t(\boldsymbol{\beta_2})$ converges uniformly in $\boldsymbol{\beta_1}, \boldsymbol{\beta_2} \in N$.

(D) $\lim\limits_{T\to\infty} \dfrac{1}{T}\sum_{t=1}^{T} [f_t(\boldsymbol{\beta_0}) - f_t(\boldsymbol{\beta})]^2 \neq 0$ if $\boldsymbol{\beta} \neq \boldsymbol{\beta_0}$.

Condition (A) follows directly from our assumption (i). To show that condition (B) is satisfied, we have from Equation (3.9) that

$$\begin{aligned}
\mathbb{P}(i \succ j|\beta_1) - \mathbb{P}(i \succ j'|\beta_2) &= \mathbb{P}(\mu_i > \mu_j|\beta_1) - \mathbb{P}(\mu_i > \mu_j'|\beta_2) && \text{(B.1)}\\
&= f_t(X_t\beta_1) - f_t(X_t\beta_2)\\
&= \int_{X_t\beta_1}^{X_t\beta_2} \frac{\partial f_t(h)}{\partial \boldsymbol{\beta}} dh\\
&= X_t(\beta_1 - \beta_2)\frac{\partial f_t(h^*)}{\partial \beta}\,, && \text{(B.2)}
\end{aligned}$$

where the last line is obtained by mean value theorem, $X_t = X_i - X_j$ and $h^* \in (X_t\beta_1, X_t\beta_2)$. Since X_t and $\partial f_t(h^*)/\partial\beta$ are all bounded, condition (B) is satisfied. Condition C is not easily verifiable, and we use Theorem 4.2.3 in Amemiya (1985). By making the assumptions (i) and (ii), we let $f_t(\boldsymbol{\beta}) = f(x_t, \boldsymbol{\beta})$ and take x_t and $f_t(\boldsymbol{\beta_1}\boldsymbol{\beta_2})$ as the y_t and $g(y_t, \theta)$ in theorem 4.2.3, respectively. We can show condition C holds in this case.

Similar to (B), the proof for Condition (D) can easily be shown using mean value theorem:

$$\begin{aligned}
\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}[f_t(\boldsymbol{\beta_0}) - f_t(\boldsymbol{\beta})]^2 &= \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}[\mathbb{P}(i \succ j|\beta_1) - \mathbb{P}(i \succ j'|\beta_2)]^2 && \text{(B.3)}\\
&= \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}[\int_{X_t\beta_1}^{X_t\beta_2} \frac{\partial f_t(h)}{\partial \boldsymbol{\beta}} dh]^2\\
&= \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}[X_t^2(\beta_1 - \beta_2)^2(\frac{\partial f_t(h^*)}{\partial \beta})^2] > 0\,. && \text{(B.4)}
\end{aligned}$$

Now, to prove asymptotic normality, apart from the assumptions above, we need to further assume:

(E) $\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\frac{\partial f_t}{\partial \boldsymbol{\beta}}\Big|_{\boldsymbol{\beta_0}}\frac{\partial f_t}{\partial \boldsymbol{\beta'}}\Big|_{\boldsymbol{\beta_0}}$ $(\equiv \boldsymbol{C})$ is a finite nonsingular matrix.

(F) $\lim_{T\to\infty}\frac{1}{T}\frac{\partial^2 S_T}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta'}}\Big|_{\boldsymbol{\beta^*}} = 2\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\frac{\partial f_t}{\partial \boldsymbol{\beta}}\Big|_{\boldsymbol{\beta_0}}\frac{\partial f_t}{\partial \boldsymbol{\beta'}}\Big|_{\boldsymbol{\beta_0}}$ whenever $\lim_{T\to\infty}\boldsymbol{\beta^*} = \boldsymbol{\beta_0}$

where $S_T = \sum_{t=1}^{T}[y_t - f_t(\boldsymbol{\beta})]^2$. Directly validating these assumptions can be difficult. Using similar idea as in Example 4.3.3 in Amemiya (1985), we can show

$$\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta_0}) = -\left[\frac{1}{T}\frac{\partial^2 S_T}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta'}}\bigg|_{\boldsymbol{\beta^*}}\right]^{-1}\frac{1}{\sqrt{T}}\frac{\partial S_T}{\partial\boldsymbol{\beta}}\bigg|_{\boldsymbol{\beta_0}}$$

where $\boldsymbol{\beta^*}$ lies between $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta_0}$. We then have

$$\begin{aligned}
\frac{1}{\sqrt{T}}\frac{\partial S_T}{\partial\boldsymbol{\beta}}\bigg|_{\boldsymbol{\beta_0}} &= -\frac{2}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_t(\frac{\partial f}{\partial\boldsymbol{\beta}}) \to N\big(0, 4\sigma_0^2\mathbb{E}[\frac{\partial f}{\partial\boldsymbol{\beta}}\frac{\partial f'}{\partial\boldsymbol{\beta}}]\big)\\
\frac{1}{T}\frac{\partial^2 S_T}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta'}}\bigg|_{\boldsymbol{\beta_0}} &= \frac{1}{T}\sum_{t=1}^{T}\big(2\frac{\partial f}{\partial\boldsymbol{\beta}}\frac{\partial f'}{\partial\boldsymbol{\beta}} + 2\epsilon_t\frac{\partial^2 f}{\partial\boldsymbol{\beta^2}}\big)\\
\lim\frac{1}{T}\frac{\partial^2 S_T}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta'}}\bigg|_{\boldsymbol{\beta_0}} &= \mathbb{E}[2\frac{\partial f}{\partial\boldsymbol{\beta}}\frac{\partial f'}{\partial\boldsymbol{\beta}}]
\end{aligned}$$

Finally we have

$$\sqrt{T}(\beta - \beta_0) \to N(0, \frac{\sigma_0^2}{\boldsymbol{C}}), \quad \boldsymbol{C} = \mathbb{E}\big[\frac{\partial f}{\partial\boldsymbol{\beta}}\frac{\partial f'}{\partial\boldsymbol{\beta}}\big]$$

For example in the case of normal distribution, $f(\cdot) \sim \mathcal{N}(0, \sigma)$, $\boldsymbol{C} = \mathbb{E}\big[\frac{1}{4\sigma^2\pi}e^{-\frac{\boldsymbol{X}\boldsymbol{\beta_0}}{2\sigma^2}}\big]$. Now it is clear that once assumptions (i)-(iii) are all satisfied, we can prove proposition 3.1 and 3.2.

Proof for Proposition 3.3. Intuitively, Proposition 3.3 can be proven by contradiction. Suppose that there exists another layout with a strictly higher value function $V' > V_{m^*}$, then it must hold that $V' > V_m$ for all m. Without loss of generality, suppose that V'_m has a cover photo of bedroom. Then, if we compare V' and V_B, there are two possible cases:

1. V' and V_B have the same cover score. Because both layouts have the same cover photo room type, the same images in the non-cover slots would have the same impact on customer utility. By construction of value functions in the algorithm, there is no other layout that would obtain a strictly higher value function for non-cover slots, which leads to a contradiction;

2. The two cover images have different quality scores. Because the algorithm selects the highest score among all bedroom photos as the cover, we have $s_{n^*} > s'_{n^*}$. However, swapping the higher quality cover image with the lower quality one cannot yield a strictly higher value function as non-cover spots have smaller impacts on a listing's attractiveness, so, again we come to a contradiction.

Appendix C

Supplemental Material for Chapter 4

C.1 Return Forecaster

An essential part of the Zalando business is the flexible return policy: for most of the countries the customers enjoy 100 day free return after the purchase. Returned articles (e.g. if they do not fit), if the pass the quality control, are coming back in the stock and can be sold later. Such a policy obviously has an impact on the pricing of the articles and plays an important part in Zalando's pricing system. In order to keep track on stock we also need to forecast returns. The return model we use consists of two components. The so called return rate model predicts the probability that a given cSKU is returned and a return times model that predicts when a given return is expected to arrive. Assuming all returns happen within a six-week window, we model return arrival by a six-dimensional vector where the i_{th} entry corresponds to the probability that a given article arrives in the $i - 1_{th}$ week after sale. The return rate models the probability of a return at a cSKU-country level. It is a decision-based model that uses two sources of information gathered over the last 52 weeks:

1. cSKU-specific information: once we observed a sufficient number of past sales, we use (observed) past returns to estimate return rates.

2. Fallback: if we do not have sufficient past sales, we use the return rate of all articles within the same article type and brand.

C.2 Proof of Corollary 1

Let us first consider a specific time slot t' and prove the corollary for the case, when inventory is sufficient to satisfy all demand, i.e. $y_{t'} \geq \sum_{c,l} z_{c,t',l} D_{c,t',l}$.

In this case in (4.8) sales become equal to demand (if corresponding discount decision is activated):

$$x_{c,t',l} = z_{c,t',l} D_{c,t',l} \quad \forall c \in [C]. \tag{C.1}$$

Which implies (assuming non-negative demand) that $\frac{x_{c,t',l}}{D_{c,t',l}} = z_{c,t',l}$, i.e. ratio between sales and demand is equal to 1 if the corresponding discount variable is chosen and zero otherwise. Let us assume without loss of generality, that such discount level is l_c for each country, i.e.

$$z_{c,t',l_c} = 1 = \frac{x_{c,t',l_c}}{D_{c,t',l_c}}. \tag{C.2}$$

At the same time, constraints (4.10)-(4.11) are forcing $\phi_{t'} = 1$:

$$\sum_c D_{c,t',l_c} = \sum_c x_{c,t',l_c} \tag{C.3}$$

$$\sum_c x_{c,t',l_c} \geq y_{t'} - M_{stock}, \tag{C.4}$$

and given that sales can never exceed demand ($D_{c,t,l} \geq x_{c,t,l}$), lead to the fact that $D_{c,t',l_c} = x_{c,t',l_c} \forall c \in [C]$.

Let us a take a look on what happens with (4.9). For every constraint, only one summand on each side of it is positive, since only one and only one $x_{c,t',l_c} > 0$ (one discount can be chosen per country). Thus the (4.9) becomes:

$$\frac{x_{c,t',l_c}}{D_{c,t',l_c}} = \frac{x_{1,t,l_c}}{D_{1,t,l_c}} \text{ for } c \in [C] \setminus \{1\},\ t \in [T], \tag{C.5}$$

which is true when $D_{c,t',l_c} = x_{c,t',l_c}$.

Let us now consider the case stock scarcity, i.e. when there is not enough stock

to satisfy all demand ($y_{t'} \leq \sum_{c,l} z_{c,t',l} D_{c,t',l}$). In this case (4.8) becomes:

$$x_{c,t',l} = z_{c,t',l} D_{c,t',l} \frac{y_{t'}}{\sum_{c,l} z_{c,t',l} D_{c,t',l}}. \tag{C.6}$$

Let us again assume that for simplicity that l_c discount level is chosen for each country c:

$$x_{c,t',l_c} = D_{c,t',l_c} \frac{y_{t'}}{\sum_{c'} D_{c',t',l_{c'}}} \tag{C.7}$$

$$x_{c,t',l} = 0 \quad \forall l \in [L] \setminus \{l_c\}, \tag{C.8}$$

which is identical (given assuming demand values) to:

$$\frac{x_{c,t',l_c}}{D_{c,t',l_c}} = \frac{y_{t'}}{\sum_{c'} D_{c',t',l_{c'}}}, \tag{C.9}$$

from which we can also deduce that $\sum_c x_{c,t',l_c} = y_{t'}$.

For constraints (4.9)-(4.11) we have:

$$\frac{x_{c,t',l_c}}{D_{c,t',l_c}} = \frac{x_{1,t',l_c}}{D_{1,t',l_c}} \qquad \text{for } \; c \in [C] \setminus \{1\}, \tag{C.10}$$

$$\sum_c D_{c,t',l_c} - x_{c,t',l_c} \leq M_{demand} \tag{C.11}$$

$$\sum_c x_{c,t,l_c} = y_{t'}, \tag{C.12}$$

where the last one comes from the fact that sales cannot exceed stock (4.4). The latter set of equations is identical to:

$$\frac{x_{c,t',l_c}}{D_{c,t',l_c}} = \frac{x_{1,t',l_c}}{D_{1,t',l_c}} \qquad \text{for } \; c \in [C] \setminus \{1\}, \tag{C.13}$$

$$\sum_c (D_{c,t',l_c} - x_{c,t',l_c}) \leq M_{demand} \tag{C.14}$$

$$\sum_c x_{c,t',l_c} = y_{t'}, \tag{C.15}$$

, from what we get

$$\frac{x_{c,t',l_c}}{D_{c,t',l_c}} = \frac{x_{1,t',l_1}}{D_{1,t',l_1}} \qquad \text{for } c \in [C] \setminus \{1\},$$
$$\sum_c x_{1,t',l_c} \frac{D_{c,t',l_c}}{D_{1,t',l_1}} = y_{t'},$$

and finally

$$x_{c,t',l_c} = D_{c,t',l_c} \frac{y_{t'}}{\sum_{c'} D_{c',t',l_{c'}}} \qquad \text{for } c \in [C] \setminus \{1\}$$

C.3 Piecewise linear approximation

The objective is to use a piecewise linear function to approximate function (4.12). Note that parametric form of the function is known, and it is continuous, twice differentiable, strictly increasing and concave. We adopt the approximation method in Kontogiorgis (2000) by selecting breakpoints on the curve, and connect them into a piecewise linear function. We adopt the infinite norm as the measure of distance for intervals of each piece,

$$\|g\|_{[a,b]} := \max_{x \in [a,b]} |g(x)| \tag{C.16}$$

which is upper bounded by $\frac{1}{8}(\Delta\tau_k)^2(\|f''\|_{[\tau_k,\tau_{k+1}]})$. Intuitively, if the function has higher curvature on a certain interval, the distance (or the approximation error) of that interval will also be larger. Therefore, we would like to put more breakpoints where the function is "more nonlinear". Formally, the paper suggests minimizing

$$\{\max_k (\Delta\tau_k)^2(\|f''\|_{[\tau_k,\tau_{k+1}]})\}$$

with breakpoints τ_k such that

$$(\Delta\tau_k)(\|f''\|_{[\tau_k,\tau_{k+1}]})^{1/2} = \text{constant}, \quad \forall k = 1, ..., K. \tag{C.17}$$

This can be approximated by

$$\int_l^{\tau_k} |f''|^{1/2} dx = \frac{k-1}{K-1} \int_l^u |f''|^{1/2} dx, \quad \forall k = 2, ..., K. \tag{C.18}$$

Algorithm 5:

initialization: Uniform breakpoints

while $\Delta\epsilon \geq \bar{\epsilon}$ **do**
  for *k = 2,...,K-1* **do**
    compute change of slope $\alpha_k := \frac{f(\tau_{k+1})-f(\tau_k)}{\tau_{k+1}-\tau_k} - \frac{f(\tau_k)-f(\tau_{k-1})}{\tau_k-\tau_{k-1}}$
    compute $\beta_k := |\alpha_k|/(\tau_{k+1} - \tau_{k-1})$
  end for
  set $\beta_1 = \beta_2$, $\beta_K = \beta_{K-1}$
  for *k = 1,...,K-1* **do**
    compute $h(x) = \beta_k + \beta_{k+1}$
  end for
  compute $G(x) := \int_l^x (h(s))^{1/2} ds$
  solve $G(\tau_k) = \frac{k-1}{K-1} G(u)$ for $k = 2, ..., K$ to get τ_k
  calculate approximation error ϵ and the change of error $\Delta\epsilon$
end while

We note that in this approach, the piecewise linear approximation function uses breakpoints **on** the original function curve, which might not be the "optimal" way of approximating function $\text{sr}(\cdot)$. Also in practice we do not have to require the breakpoints to be on the curve, therefore there could be potential improvements.

Bibliography

Amemiya T (1985) *Advanced econometrics* (Harvard university press).

Anderson SP, De Palma A, Thisse JF (1992) *Discrete choice theory of product differentiation* (MIT press).

Aouad A, Farias VF, Levi R (2015) Assortment optimization under consider-then-choose choice models. *Working Paper, Available at SSRN 2618823* .

Aouad A, Feldman J, Segev D, Zhang D (2019) Click-based mnl: Algorithmic frameworks for modeling click data in assortment optimization. *Available at SSRN 3340620* .

Ban GY, Keskin NB (2020) Personalized dynamic pricing with machine learning: High dimensional features and heterogeneous elasticity. *Available at SSRN 2972985* .

Bateni MH, Chen Y, Ciocan D, Mirrokni V (2018) Fair resource allocation in a volatile marketplace. *Available at SSRN 2789380* .

Ben-Akiva ME, Lerman SR (1985) *Discrete choice analysis: theory and application to travel demand*, volume 9 (MIT press, Cambridge, MA).

Bernstein F, Kök AG, Xie L (2015) Dynamic assortment customization with limited inventories. *Manufacturing & Service Operations Management* 17(4):538–553.

Bernstein F, Martínez-de Albéniz V (2017) Dynamic product rotation in the presence of strategic customers. *Management Science* 63(7):2092–2107.

Berthene A (2022) *Coronavirus pandemic adds $219 billion to US ecommerce sales in 2020-2021.* `https://www.digitalcommerce360.com/article/coronavirus-impact-online-retail/` [Accessed: 2022-03-21].

Bertsimas D, Farias VF, Trichakis N (2011) The price of fairness. *Operations research* 59(1):17–31.

Bertsimas D, Farias VF, Trichakis N (2012) On the efficiency-fairness trade-off. *Management Science* 58(12):2234–2250.

Besbes O, Sauré D (2016) Product assortment and price competition under multinomial logit demand. *Production and Operations Management* 25(1):114–127.

Boada-Collado P, Martínez-de Albéniz V (2020) Estimating and optimizing the impact of inventory on consumer choices in a fashion retail setting. *Manufacturing & Service Operations Management* 22(3):582–597.

Bostoen F (2018) Neutrality, fairness or freedom? principles for platform regulation. *Principles for Platform Regulation (March 31, 2018). Internet Policy Review* 7(1):1–19.

Bu J, Simchi-Levi D, Xu Y (2020) Online pricing with offline data: Phase transition and inverse square law. *International Conference on Machine Learning*, 1202–1210 (PMLR).

Cachon GP, Netessine S (2006) Game theory in supply chain analysis. *Models, methods, and applications for innovative decision making* 200–233.

Caro F, Gallien J (2012) Clearance pricing optimization for a fast-fashion retailer. *Operations Research* 60(6):1404–1422.

Caro F, Martínez-de Albéniz V (2012) Product and price competition with satiation effects. *Management Science* 58(7):1357–1373.

Caro F, Martínez-de Albéniz V (2015) Fast fashion: Business model overview and research opportunities. *Retail supply chain management*, 237–264 (Springer).

Caro F, Martínez-de Albéniz V, Rusmevichientong P (2014) The assortment packing problem: Multiperiod assortment planning for short-lived products. *Management Science* 60(11):2701–2721.

Cheung WC, Simchi-Levi D, Wang H (2017) Dynamic pricing and demand learning with limited price experimentation. *Operations Research* 65(6):1722–1731.

Cohen M, Elmachtoub AN, Lei X (2019) Price discrimination with fairness constraints. *Available at SSRN 3459289* .

Cui R, Li J, Zhang DJ (2019) Reducing discrimination with reviews in the sharing economy: Evidence from field experiments on airbnb. *Management Science, Forthcoming* .

Datta R, Joshi D, Li J, Wang JZ (2006) Studying aesthetics in photographic images using a computational approach. *European conference on computer vision*, 288–301 (Springer).

den Boer A, Keskin NB (2019) Dynamic pricing with demand learning and reference effects. *Available at SSRN 3092745* .

den Boer AV (2015) Dynamic pricing and learning: historical origins, current research, and new directions. *Surveys in operations research and management science* 20(1):1–18.

Desrosiers J, Lübbecke ME (2005) A primer in column generation. *Column generation*, 1–32 (Springer).

Dzyabura D, Jagabathula S (2018) Offline assortment optimization in the presence of an online channel. *Management Science* 64(6):2767–2786.

Farias VF, Jagabathula S, Shah D (2013) A nonparametric approach to modeling choice with limited data. *Management science* 59(2):305–322.

Farronato C, Fradkin A (2018) The welfare effects of peer entry in the accommodation market: The case of airbnb. Technical report, National Bureau of Economic Research.

Federgruen A, Hu M (2015) Multi-product price and assortment competition. *Operations Research* 63(3):572–584.

Feldman J, Zhang D, Liu X, Zhang N (2019) Customer choice models versus machine learning: Finding optimal product displays on alibaba. *Available at SSRN* .

Ferreira KJ, Goh J (2021) Assortment rotation and the value of concealment. *Management Science* 67(3):1489–1507.

Ferreira KJ, Lee BHA, Simchi-Levi D (2016) Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & Service Operations Management* 18(1):69–88.

Fisher M, Gallino S, Li J (2018) Competition-based dynamic pricing in online retailing: A methodology validated with field experiments. *Management science* 64(6):2496–2514.

Fox JT (2007) Semiparametric estimation of multinomial discrete-choice models using a subset of choices. *The RAND Journal of Economics* 38(4):1002–1019.

Fradkin A (2017) Search, matching, and the role of digital marketplace design in enabling trade: Evidence from airbnb. *Working Paper, Boston University* .

Frangioni A, Gendron B, Gorgone E (2015) *On the computational efficiency of subgradient methods: a case study in combinatorial optimization* (CIRRELT, Centre interuniversitaire de recherche sur les réseaux d'entreprise ...).

Fudenberg D, Tirole J (1991) Game theory, 1991. *Cambridge, Massachusetts* 393(12):80.

Gallego G, Wang R (2014) Multiproduct price optimization and competition under the nested logit model with product-differentiated price sensitivities. *Operations Research* 62(2):450–461.

Gorn GJ, Chattopadhyay A, Yi T, Dahl DW (1997) Effects of color as an executional cue in advertising: They're in the shade. *Management science* 43(10):1387–1400.

He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Heese HS, Martínez-de Albéniz V (2018) Effects of assortment breadth announcements on manufacturer competition. *Manufacturing & Service Operations Management* 20(2):302–316.

Jagabathula S, Mitrofanov D, Vulcano G (2019) Inferring consideration sets from sales transaction data. *Available at SSRN 3410019* .

Kanoria Y, Saban D (2020) Facilitating the search for partners on matching platforms. *Management Science (to appear)* .

Kelley JE Jr (1960) The cutting-plane method for solving convex programs. *Journal of the society for Industrial and Applied Mathematics* 8(4):703–712.

Keskin NB, Zeevi A (2014) Dynamic pricing with an unknown demand model: Asymptotically optimal semi-myopic policies. *Operations Research* 62(5):1142–1167.

Koetsier J (2020) *COVID-19 Accelerated E-Commerce Growth '4 To 6 Years'*. `https://www.forbes.com/sites/johnkoetsier/2020/06/12/covid-19-accelerated-e-commerce-growth-4-to-6-years/?sh=693478a600fa` [Accessed: 2022-03-21].

Kök AG, Fisher ML (2007) Demand estimation and assortment optimization under substitution: Methodology and application. *Operations Research* 55(6):1001–1021.

Kök AG, Fisher ML, Vaidyanathan R (2008) Assortment planning: Review of literature and industry practice. *Retail supply chain management* 99–153.

Kontogiorgis S (2000) Practical piecewise-linear approximation for monotropic optimization. *INFORMS Journal on Computing* 12(4):324–340.

Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 1097–1105.

Lei Y, Jasin S, Uichanco J, Vakhutinsky A (2021) Joint product framing (display, ranking, pricing) and order fulfillment under the multinomial logit model for e-commerce retailers. *Manufacturing & Service Operations Management* .

Li H, Simchi-Levi D, Wu MX, Zhu W (2019) Estimating and exploiting the impact of photo layout in the sharing economy. *Available at SSRN 3470877* .

Li J, Moreno A, Zhang DJ (2016) Pros vs joes: Agent pricing behavior in the sharing economy. *Working Paper, University of Michigan* .

Ma W, Simchi-Levi D, Zhao J (2018) Dynamic pricing under a static calendar. *arXiv preprint arXiv:1811.01077* .

Mahajan S, Van Ryzin G (2001) Stocking retail assortments under dynamic consumer substitution. *Operations Research* 49(3):334–351.

Manski CF (1975) Maximum score estimation of the stochastic utility model of choice. *Journal of econometrics* 3(3):205–228.

Martinez de Albeniz V (2017) The production department at privalia. *Case, IESE Business School, Barcelona, Spain* .

Martínez-de Albéniz V, Roels G (2011) Competing for shelf space. *Production and Operations Management* 20(1):32–46.

Martinez-de Albeniz V, Valdivia A (2019) Measuring and exploiting the impact of exhibition scheduling on museum attendance. *Manufacturing & Service Operations Management* 21(4):761–779.

McCoy JH, Lee HL (2014) Using fairness models to improve equity in health delivery fleet management. *Production and Operations Management* 23(6):965–977.

McFadden D (1978) Modeling the choice of residential location. *Transportation Research Record* (673).

Meyers-Levy J, Peracchio LA (1992) Getting an angle in advertising: The effect of camera angle on product evaluations. *Journal of marketing research* 29(4):454–461.

Mikels JA, Fredrickson BL, Larkin GR, Lindberg CM, Maglio SJ, Reuter-Lorenz PA (2005) Emotional category data on images from the international affective picture system. *Behavior research methods* 37(4):626–630.

Miller EG, Kahn BE (2005) Shades of meaning: the effect of color and flavor names on consumer choice. *Journal of consumer research* 32(1):86–92.

Mo J, Walrand J (2000) Fair end-to-end window-based congestion control. *IEEE/ACM Transactions on networking* 8(5):556–567.

N Boute R, Gijsbrechts J, A Van Mieghem J (2020) Digital lean operations: Smart automation and artificial intelligence in financial services. Babich V, Birge J, Hilary G, eds., *Innovative Technology at the interface of Finance and Operations. Springer Series in Supply Chain Management* (Springer Nature), forthcoming.

Nambiar M, Simchi-Levi D, Wang H (2019) Dynamic learning and pricing with model misspecification. *Management Science* 65(11):4980–5000.

Natter M, Reutterer T, Mild A, Taudes A (2007) Practice prize report—an assortmentwide decision-support system for dynamic pricing and promotion planning in diy retailing. *Marketing Science* 26(4):576–583.

Özer Ö, Ozer O, Phillips R (2012) *The Oxford handbook of pricing management* (Oxford University Press).

Papke LE, Wooldridge JM (1996) Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. *Journal of applied econometrics* 11(6):619–632.

Qiang S, Bayati M (2016) Dynamic pricing with demand covariates. *Available at SSRN 2765257* .

Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* .

Smith SA, Achabal DD (1998) Clearance pricing and inventory policies for retail chains. *Management Science* 44(3):285–300.

Snyder M, DeBono KG (1985) Appeals to image and claims about quality: Understanding the psychology of advertising. *Journal of personality and Social Psychology* 49(3):586.

Talluri K, Van Ryzin G (2004) Revenue management under a general discrete choice model of consumer behavior. *Management Science* 50(1):15–33.

Talluri KT, Van Ryzin GJ (2006) *The theory and practice of revenue management*, volume 68 (Springer Science & Business Media).

Valdez P, Mehrabian A (1994) Effects of color on emotions. *Journal of experimental psychology: General* 123(4):394.

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. *Advances in neural information processing systems*, 5998–6008.

Vulcano G, Van Ryzin G, Chaar W (2010) Om practice - choice-based revenue management: An empirical study of estimation and optimization. *Manufacturing & Service Operations Management* 12(3):371–392.

Wagner L, Martínez-de Albéniz V (2020) Pricing and assortment strategies with product exchanges. *Operations Research* 68(2):453–466.

Wang R (2012) Capacitated assortment and price optimization under the multinomial logit model. *Operations Research Letters* 40(6):492–497.

Wang R (2018) When prospect theory meets consumer choice models: Assortment and pricing management with reference prices. *Manufacturing & Service Operations Management* 20(3):583–600.

Zhang S, Lee D, Singh PV, Srinivasan K (2017) How much is an image worth? airbnb property demand estimation leveraging large scale image analytics. *Working Paper, Carnegie Mellon University* .

Zhang S, Mehta N, Singh PV, Srinivasan K (2019) Can lower-quality images lead to greater demand on airbnb? *Working Paper, Carnegie Mellon University* .