EL COLEGIO DE MÉXICO
CENTRO DE ESTUDIOS ECONÓMICOS

MAESTRÍA EN ECONOMÍA

TRABAJO DE INVESTIGACIÓN PARA OBTENER EL GRADO DE MAESTRO EN ECONOMÍA

ESTIMATING A MODEL OF DYNAMIC PRICING WITH MENU COSTS FOR AIRBNB IN MEXICO CITY

JULIO CÉSAR PASTOR HERNÁNDEZ

PROMOCIÓN 2022-2024

ASESOR:

DR. EDWIN ALEXANDER MUÑOZ RODRÍGUEZ

JULIO 2024

This research was partially funded by the Ford Foundation through a grant awarded to the PRAEM program at El Colegio de Mexico A.C. The views and conclusions expressed in this dissertation are solely those of the authors and do not necessarily reflect those of the Ford Foundation.

Abstract

This study examines dynamic pricing in the Airbnb market of Mexico City, focusing on the costs hosts incur when adjusting prices over time. Using a discrete choice model of demand for accommodations and a dynamic discrete choice model for the pricing decision, we find that menu costs in Mexico City are significantly high, often exceeding 300% of the average rental price, and only decreasing to 87% near the arrival date. The price adjustment costs remain particularly high throughout the selling period, with a significant decrease only when the lead time is zero. This suggests that hosts are only willing to change their prices when the arrival date is very close, generating substantial price stickiness in this market. Future research should investigate menu costs between professional and non-professional hosts, address price endogeneity, and conduct counterfactual analysis to policy evaluation.

Life can only be understood going backwards, but it must be lived going forwards.

Søren Kierkegaard

To Nayeli, for your unconditional love and support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

According to Den Boer (2015), dynamic pricing involves determining optimal selling prices for products or services in settings where prices can be frequently and easily adjusted. This approach is applicable to both online vendors and physical stores using digital price tags. Digital technology enables continuous price adjustments in response to changing circumstances without incurring costs or effort. Dynamic pricing is widely adopted across various industries and is often considered an essential component of pricing strategies. It involves adaptive policies that account for competitor pricing, enhance customer segmentation, and adjust to market conditions and time sensitivity (when the items are perishable).

Focusing specifically on hosts who offer their listings on Airbnb, the expected rational behavior for making optimal intertemporal decisions can be hindered by various factors. For instance, if income from the platform is not their primary source of income, hosts might find the administrative task of continuously managing listing prices burdensome and thus may avoid adjusting them. Additionally, another challenge in setting the optimal price is the need to understand not only the market value of the listing, but also the expected future demand. All factors that prevent a host from changing their prices can be summarized as adjustment costs, commonly referred to in the literature as *menu costs*.

High menu costs in the temporary lodging industry may cause price rigidity, limiting hosts' ability to adjust prices according to market conditions. This leads to market inefficiencies and a competitive disadvantage for hosts who cannot implement dynamic pricing. Furthermore, it increases administrative burdens, discourages market entry, and may negatively affect customer satisfaction due to perceived unfair pricing.

With this in mind, this study aims to be pioneer in quantifying the menu costs faced by Airbnb hosts in Mexico City for the first time. This will enhance our understanding of the price dynamics in this market. Precisely, this research work seeks to answer the following questions: (i) What proportion do menu costs represent relative to the average rental price set in Mexico City?, (ii) Do these costs exhibit any particular behavior as the arrival date approaches?, and (iii) How do these costs compare with findings in other cities?

To answer these questions, we estimate a dynamic pricing model that includes price adjustment costs. This specification is inspired by the one proposed by Pan and Wang (2021) and provides an econometric framework for analyzing the Airbnb market from both the demand and supply sides. On one hand, the model assumes consumers arrive randomly, with the probability of a listing being in demand depending on both the consumer arrival rate and the choice probability given by a logit model. On the other hand, the host's pricing problem is framed as a dynamic programming problem, incorporating menu costs into the intertemporal optimization process.

To improve the estimation process, (i) listing attributes considered for estimating demand probabilities were reduced using variable clustering, and (ii) listings were condensed using the K-means method to synthesize the number of competitors in the choice probabilities computation. Since the pricing problem is host-specific, the latter approach not only helps in reducing dimension for estimating demand parameters, but also reduces the number of dynamic programming problems that need to be solved. The estimation process is divided into two stages and employs the maximum likelihood method.

The main finding of this research for Mexico City revealed that adjustment costs are significantly high relative to the average rental price, exceeding 300% of its value. They only decrease substantially when the arrival date is very close, reaching 87% of the average rental price. We can compare these percentages with those found in Manhattan by Pan and Wang (2021), which are less than 3%. This observation suggests that hosts in Mexico City generally face high menu costs, resulting in suboptimal pricing strategies.

Our estimates are limited by the lack of granular pricing data. Thus, future research should use more detailed data to verify the robustness of our findings. Moreover, related studies have shown that professional and non-professional hosts employ different pricing strategies (Abrate et al. (2022)). Further research on this issue is granted.

The rest of the study is structured as follows: Section 2 reviews the related literature. Section 3 introduces the sampled data used to fit the structural model specified in Section 4. Section 5 describes the estimation process and discusses the results obtained for Mexico City. Finally, Section 6 presents the conclusions and suggests directions for future research.

Chapter 2

Literature review

Den Boer (2015) provided a comprehensive literature review on dynamic pricing and learning, in which the author points out that in recent years there has been a rapid expansion of literature in this field, with contributions from diverse scientific communities: operations research and management sciences, marketing, computer science, economics, and econometrics. The goal of this chapter is to bolster the existing literature review and align it with the most current studies in the application of dynamic pricing in the lodging industry.

Since Airbnb is a market for accommodations on specific arrival dates, it is by definition a market of perishable items. With this in mind, we will begin with a brief revision on this type of markets. Lazarev (2011) studied how the ability to price discriminate over time affects production, product quality, and product allocation among customers. The theoretical model considers forward-looking heterogeneous consumers and a monopoly firm who can affect the quality and quantity of the goods sold each period. Prices were observed daily but bookings were only observed quarterly. In this study, intertemporal price discrimination lowers ticket quality and prices for leisure travelers, but raises prices and reduces supply for business travelers. Resale benefits business travelers and improves short-term social welfare but may reduce airline profits and lead to market exit, while still achieving over 90% of the profits from third-degree price discrimination.

Escobari (2012) and Alderighi et al. (2015) identified evidence indicating that airlines deal with stochastic demand and that prices are adjusted based on in-flight seat availability and purchasing date. On the one hand, Escobari (2012) found that price increases as the inventory decreases, and decreases as the flight date nears. The author points out that these findings are consistent with various theoretical models of optimal pricing under uncertain demand and perishable inventories (e.g., Prescott (1975), Gallego and Van Ryzin (1994), Zhao and Zheng (2000), and Su (2007)). On the other hand, Alderighi et al. (2015) mentioned that their study supports the findings in Gerardi and Shapiro (2009) that the lack of competitive pressure allows airlines to extract more surplus from consumers with more inelastic demand.

Williams (2022) evaluated the welfare effects of dynamic airline pricing proposing a model that allows intertemporal price discrimination and dynamic adjustment to stochastic demand, finding that firm revenue and consumer welfare could be improved in comparison with a uniform pricing strategy. Leisure consumers benefit from dynamic pricing, meanwhile the airline could optimally use fares as a response to demand shocks for business consumers.

This study is directly related to research on dynamic pricing in Airbnb using structural models, representing another approach to this issue.[1] Pan and Wang (2021) examined the revenue and welfare effects of implementing automated pricing for Airbnb listings in New York City. The model focused on factors such as price adjustment costs, fluctuating willingness to pay among customers, inventory structure, and competition among sellers. The main findings were that (i) price rigidity could be rationalized by a price adjustment cost between 0.9-2.2% of the listed price; and (ii) automated pricing has the potential to boost revenue for both the platform and hosts by 4.8% and 3.9%, respectively; customer welfare resulted inconclusive upon the duration of their stays. This research will follow the model proposed by Pan and Wang (2021).

Huang (2022) studied Airbnb listings in San Francisco, California; data was obtained from Inside Airbnb website. It was showed that there are prevalent pricing frictions on the platform, and these frictions provoke a 14% consumer welfare loss and a 0-15% seller-profit loss. Price-setting costs and cognitive constraints are plausible explanations of the frictions. To eliminate almost all frictions, author proposed a dynamic programming model that explains how prices should vary by

[1] Extensive research on the platform employs descriptive techniques to study dynamic pricing. For more details, refer to Abrate et al. (2022).

market conditions, letting sellers decide on the price levels.

More broadly, we applied techniques from the literature on dynamic discrete choice, a field initiated by Rust (1987). The author studied the optimal timing of bus engine replacement at the Madison (Wisconsin) Metropolitan Bus Company providing an optimal stopping rule which is the solution to a stochastic dynamic programming problem.

Notably, studies Eckstein and Wolpin (1989), Rust (1994), Aguirregabiria and Mira (2010), and Arcidiacono and Ellickson (2011) serve as reference surveys and are central to the literature in structural econometrics. The methods detailed in those studies have been applied to analyze the lodging industry. Cho and Rust (2018) analyzed a confidential reservation database provided by a luxury hotel, where setting prices in this industry is a challenging high-dimensional problem since hotels must not only quote prices for current and future dates, but also deal with different types of rooms and customers. They found that a simple price-following strategy is suboptimal in comparison with their dynamic programming approach, where authors concluded that pricing strategy of the hotel they studied is competitive and is best described as a rational best response to its beliefs about demand and the prices set by its competitors. Merlo et al. (2015) formulated and solved a discrete-time, finite-horizon dynamic programming problem of the seller's optimal strategy for selling a house. The model is able to capture important features of the data, including the relatively high degree of stickiness of listing prices; authors found that a tiny menu cost of changing the listing prices is sufficient to explain this fact.

For studies applied to Mexico City, we find López and Ramírez-Álvarez (2021) and Merino and Muñoz-Rodríguez (2024). On the one hand, López and Ramírez-Álvarez (2021) implemented a hedonic pricing model to explain the differences in price levels and the marginal contribution of attributes on the nightly price for listings in Mexico City. In addition to listing features, crime levels, access to public transportation routes, and sentiment analysis on reviews were included as explanatory variables. The main findings were that (i) the null hypothesis that professional hosts have an advantage when setting their prices over non-professional hosts was not rejected; setting a price 16.2% higher for a similar listing; and (ii) the crime incidence rate and access to public transportation routes affect negatively nightly prices. On the other hand, Merino and Muñoz-Rodríguez (2024) estimated a daily supply and demand model for Airbnb in Cuauhtémoc, Mexico City; the demand was characterized with a discrete choice model with random coefficients with endogeneity and the supply was assumed to be price competition by hosts. The primary conclusions were that (i) hosts have a low Lerner index (0.18, annual average) indicating that they have little influence on pricing; and (ii) no significant discrepancy in market power was observed when categorizing hosts by type (professional vs. non-professional). Although both López and Ramírez-Álvarez (2021) and Merino and Muñoz-Rodríguez (2024) concluded that there is a greater willingness to pay for listings offered by professional hosts, they differ in the conclusion regarding the market power of these hosts in setting prices, possibly because Merino and Muñoz-Rodríguez (2024) only analyzed the Cuauhtémoc borough.

Despite Airbnb's rapid expansion in Mexico, there are limited literature specifically addressing this market. In addition to López and Ramírez-Álvarez (2021) and Merino and Muñoz-Rodríguez (2024), de Oca et al. (2018) and Ruiz-Correa et al. (2019) have examined the topic of Airbnb in Mexico City from an Urban Studies standpoint. Banco de México (2021) explored the progression of the Airbnb market in Mexico City. To our best knowledge, this research work is the first to analyze the price dynamics of Airbnb in Mexico City estimating a model of dynamic pricing; our aim is to delve deeper into the existing literature on this market, covering some of the recommendations made by López and Ramírez-Álvarez (2021) (implementing a dynamic model and capturing seasonal effects) and Merino and Muñoz-Rodríguez (2024) (introducing a supply model with dynamic optimization).

Other studies further removed from ours examine the optimality of pricing strategies in various industries, comparing them to those derived from a dynamic programming model. In this direction, Cho and Rust (2010) studied the car rental market, concluding that rental markets are not fully competitive and firms may be behaving suboptimally. Authors provided a case study of a large car rental company and suggested a properly chosen declining rental price function to increase overall revenues. The results from an experiment were consistent with the predictions of the model, which estimated an improvement of 6-140% in profitability (depending on the vehicle type).

Misra and Nair (2011) analyzed sales-force compensation schemes for a large contact lens manufacturer, applying dynamic programming-based solutions combined with structural specifications of worker behavior. Recommendations of the study were then implemented at the firm, resulting in a 9% improvement in overall revenues (translated to about $12 million incremental revenues annually). McClelland and Rust (2018) examined optimal timing of replacement in the equipment rental industry. The benefits of the optimal replacement strategy proposed by authors arise from leveraging the seasonal fluctuations in rental demand and the timing of economic cycles; for some machines, the optimal strategy result to be procyclical, but for some others, countercyclical.

To the best of our knowledge, our research contributes to the existing literature in several ways: (i) we achieved the first dynamic pricing model incorporating menu costs specifically tailored to Airbnb operations in Latin America. Moreover, we addressed recommendations suggested by López and Ramírez-Álvarez (2021) and Merino and Muñoz-Rodríguez (2024); and (ii) our estimated model offers the potential for conducting counterfactual analyses to assess the effects of reducing the price adjustment costs on the equilibrium of host pricing behavior. This approach provides valuable insights into the social welfare of facilitating price-setting processes within the platform.

Chapter 3

Background and data

3.1 Background

Huang (2022) excelled in presenting the most relevant features about Airbnb. According to this author, Airbnb holds a dominant position in the short-term rental market. Individuals who offer accommodations, known as “hosts”, list their properties on the platform and establish prices. Potential renters, or “guests”, access the platform’s app/website to search for properties by destination city and check-in/check-out dates, and then proceed to make booking decisions. When a booking is confirmed, guests pay the nightly rate, a cleaning fee per booking (both determined by the host), a service fee (set by the platform), and applicable taxes. Generally, guests pay a 14% service fee, meanwhile hosts pay a 3% platform fee.[1] In Mexico City, guests pay a 3-5% transient lodging tax. [2]

Author argued that pricing frictions could stem from Airbnb’s pricing interface: within this interface, hosts can establish a single base price for all nights and a separate weekend rate for Fridays and Saturdays. To exemplify this, figure 3.1 shows the average rental prices in Mexico City for arrival dates in the first quarter of 2024; a weekly seasonal behavior can be observed for weekends. Additionally, hosts have the option to specify nightly rates on a price calendar (selecting a

[1] https://www.airbnb.com/resources/hosting-homes/a/how-much-does-airbnb-charge-hosts-288?locale=en&_set_bev_on_new_domain=1716689816_Zjc1NjhkNGI1OGNm. Accessed in May 2024.

[2] https://www.airbnb.com/help/article/2288?locale=en&_set_bev_on_new_domain=1716689816_Zjc1NjhkNGI1OGNm. Accessed in May 2024.

period of time). Once a nightly rate is configured, the only way to change it is through manual adjustment, so it represents a labor-intensive task.

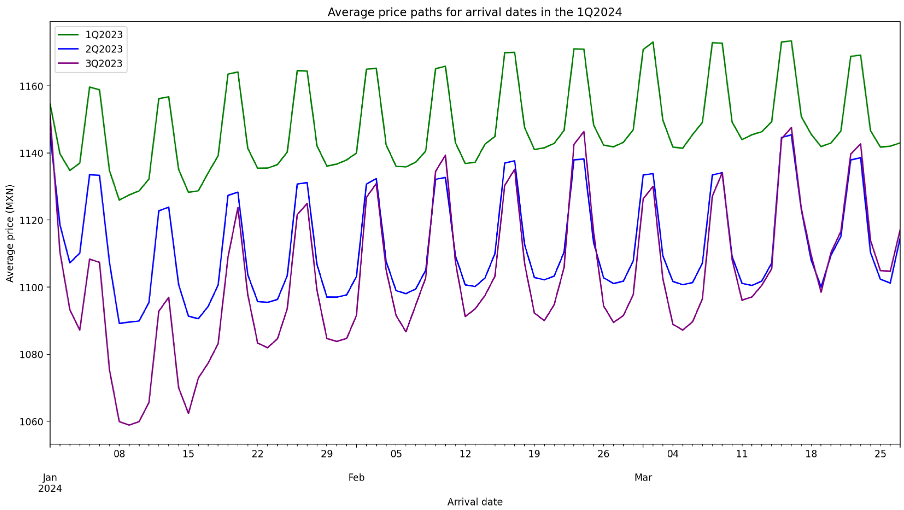

Figure 3.1: Average rental prices in Mexico City for arrival dates in the 1Q2024.

Huang (2022) also detailed two modifications introduced by Airbnb that impact price flexibility. Firstly, in November 2015, Airbnb introduced a pricing algorithm called “Smart Pricing”, enabling automatic adjustments of nightly prices based on demand, with the option to set minimum and maximum price thresholds. Ye et al. (2018), who are affiliated with Airbnb, explained that (one deployed in production version of) the algorithm estimates a simplified consumer demand function by utilizing observed prices, and then determines the prices that maximize revenue for each listing.

Secondly, the debut of “Last-Minute Discounts” in early 2019. This feature allows hosts to set a percentage discount on prices and a lead time threshold (the time before arrival date). If the lead time falls below the threshold, the percentage discount is automatically applied to the nightly rate, eliminating the need for manual intervention by the host. This feature is only a part of the “professional hosting tools” offered by Airbnb.

3.2 Data

Data was downloaded from Inside Airbnb [1] under the Creative Commons Attribution 4.0 International License.[2] The website encompasses all listings across various cities and typically gathers data from Airbnb once per month. However, for data related to Mexico City, the current frequency of data collection is quarterly. For this study, data from the first three quarters of 2023 will be used; the decision was made to exclude the final quarter of this year from the analysis due to observed inconsistencies in price data. Price information for the first quarter of 2024 (lead time $t = 0$) was obtained from AirDNA through my advisor. The integration of Inside Airbnb and AirDNA data was accomplished using the (numeric) listing ID. For further details regarding the data description about Inside Airbnb files, reference can be made to the data dictionary.[3]

Despite all the available datasets, only the "listings.csv" and "calendar.csv" files are relevant to us. The first dataset includes listing characteristics on each sampling quarter. Starting from sampling date t, the second dataset comprises the availability status and established price for each night throughout the subsequent calendar year, for every listing. With this in mind, we have up to four quarterly observations for each night in the future. Figure 3.2 shows the geographical location of listings on Airbnb in Mexico City in 4Q2023. [4]

3.2.1 Sample selection

3.2.1.1 Censorship and truncation

In both the listing and calendar datasets, there are listings where information for all quarters of 2023 is unavailable. To ensure consistency, listings were selected where data is provided throughout the entire year.

Similarly, pricing and availability data are censored, as we only have data from

[1] `https://insideairbnb.com/`

[2] `https://creativecommons.org/licenses/by/4.0/`

[3] `https://docs.google.com/spreadsheets/d/1iWCNJcSutYqpULSQHlNyGInUvHg2BoUGoNRIGa6Szc4/edit#gid=1322284596`. Accessed in April 2024.

[4] Source: `https://insideairbnb.com/mexico-city`. Accessed in April 2024.

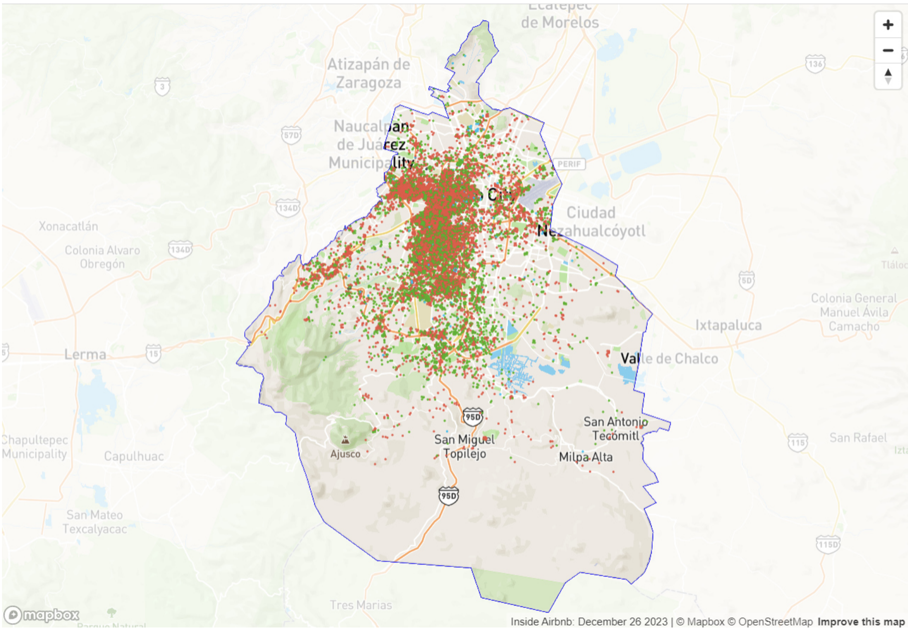

Figure 3.2: Geographical location of listings on Airbnb in Mexico City in 4Q2023.

the dates when web scraping was conducted by Inside Airbnb. This study does not employ any techniques for interpolating or extrapolating censored information.

3.2.1.2 Variable clustering

Variable clustering is applied to evaluate collinearity, redundancy, and to group variables into clusters that can be considered as a single variable, thus assisting in data reduction. We gathered listing data for the first quarter of 2023 and performed hierarchical cluster analysis on variables, using the Spearman correlation as the similarity measure.

Figure 3.3 displays the correlation among variables organized into clusters. The 33 variables under analysis consist of numerical values, with some being continuous and others categorical ordinal. It is noteworthy that for variables with missing values, no imputation was conducted; instead, these variables were categorized, and an additional category was introduced to denote the presence of missing values.

In the initial cluster depicted in figure 3.3, we have assembled variables related to the quantity of bathrooms, bedrooms, and beds within the listing, alongside

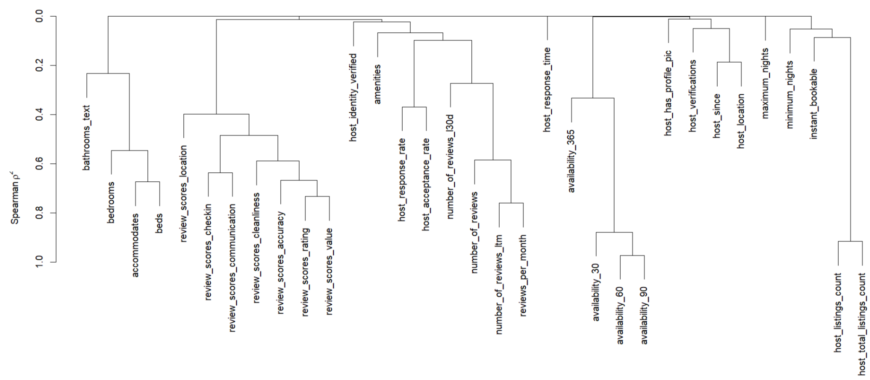

Figure 3.3: Variable clustering for numerical variables on listings dataset.

the maximum capacity for accommodates. This cluster exemplifies how these variables collectively indicate the size of the listing, thereby suggesting that solely the "accommodates" attribute could serve as a representative of the cluster. Figure 3.4 illustrates the selected attributes to be used in the model. Whereas correlation values in figure 3.3 reached up to 1, in figure 3.4 these levels decreased to a maximum of 0.35. Table A.1 provides a concise overview of the variables utilized in the model.

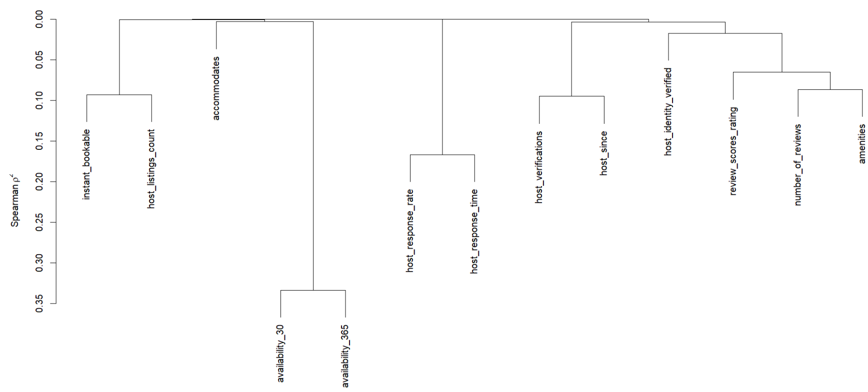

Figure 3.4: Variable clustering for selected listings dataset attributes to be used in the model.

3.2.1.3 Listing clustering

Since shared and hotel rooms account for only about 1.7% of the listings in Mexico City, a first subsample was taken excluding them. In figure 3.2, private rooms are represented by the green dots while entire home/apartment are represented in red; the proportion of these listings is 33.3% and 65.1%, respectively. Since our model does not consider cancellations during the reservation period, a second subsample was obtained after excluding all listings that had presented at least one cancellation in calendar data.

Following the ideas proposed by Pan and Wang (2021) and Huang (2022), we clustered all subsampled listings into segments based on their observed characteristics (described in table A.1) defining subgroups at neighbourhood, superhost indicator, and room type level; listing prices were not considered in the clustering process. Listing data used for profiling corresponded to the first quarter of 2023. Each defined subgroup was limited to a maximum of three clusters, determined by optimizing the number of clusters using the K-means method and considering the Silhouette criterion. Different subgroups might have different numbers of clusters.

Figures 3.5, 3.6, 3.7, and 3.8 are merely illustrative examples of the clusters identified. They depict representations in a two-dimensional space of sets that exist in a 12-dimensional space. Thus, the Principal Component Analysis (PCA) method captures approximately 30% of the total variance with two variables. PCA was solely employed for projection purposes to aid visualization; no artificial variables were generated from the listing characteristics. In the model outlined in section 4.2, only the attributes detailed in table A.1 are considered.

With a total of 142 clusters in Mexico City, there exist an equivalent number of centroids, which together represent all the listings in the city. These centroids, while not necessarily part of the original dataset, were obtained from the defined subgroups minimizing the distance from the theoretical centroid. The Euclidean metric was used for this purpose.

Considering all the details provided earlier, the reduced listing dataset comprises 142 listings, each characterized by 16 attributes. Additionally, the reduced calendar dataset encompasses the availability and pricing details of these 142 listings for arrival dates from January 1$^{\text{st}}$ to February 29$^{\text{th}}$, 2024. Hence, the data

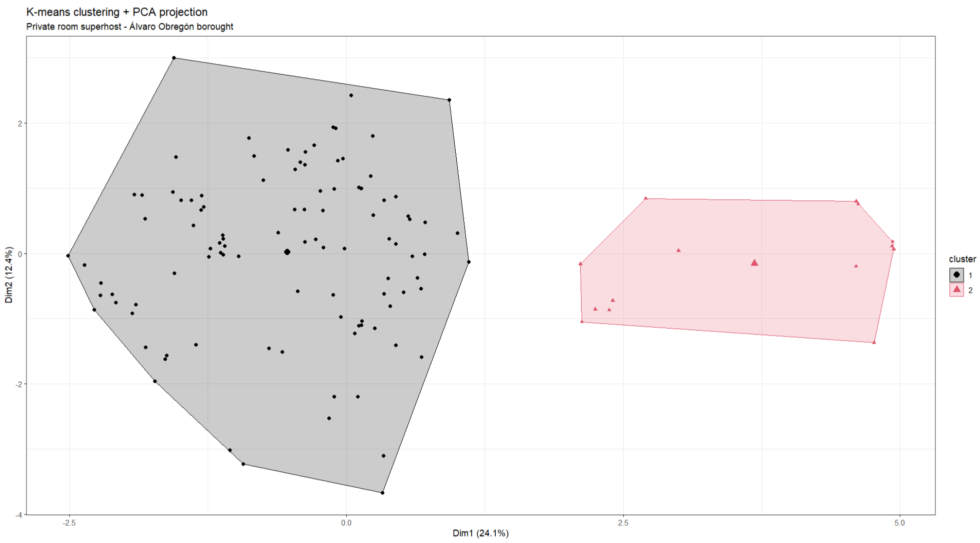

Figure 3.5: Clusters identified for superhosts offering private rooms in Álvaro Obregón.

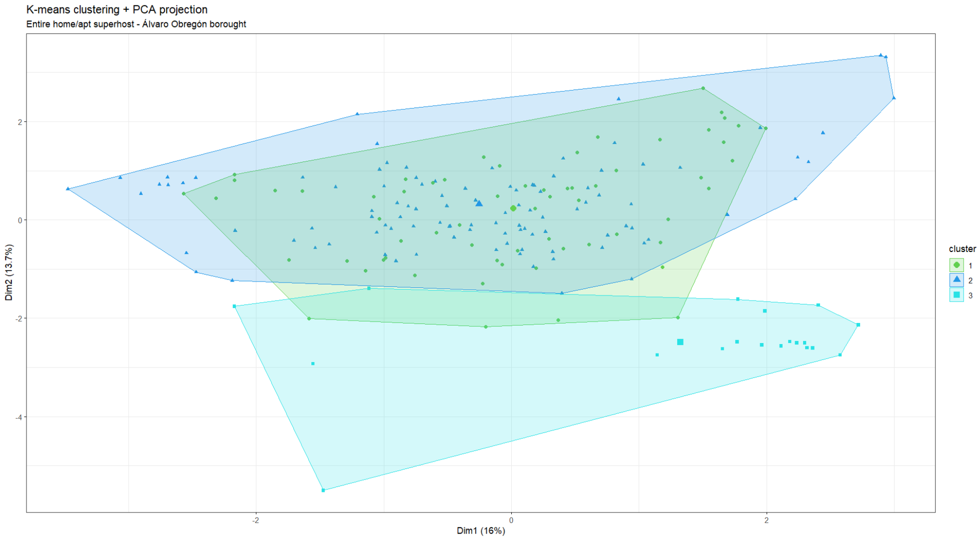

Figure 3.6: Clusters identified for superhosts offering entire home/apt in Álvaro Obregón.

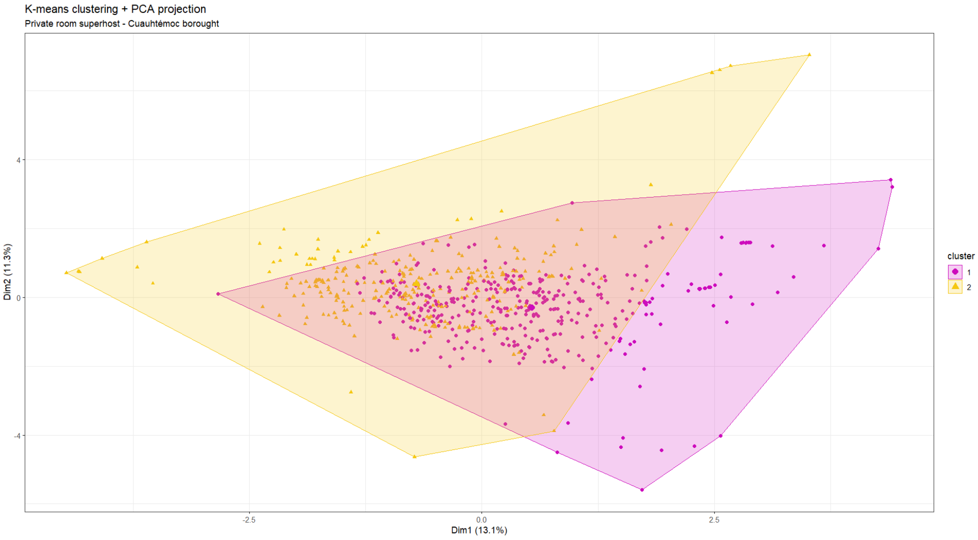

Figure 3.7: Clusters identified for superhosts offering private rooms in Cuauhtémoc.

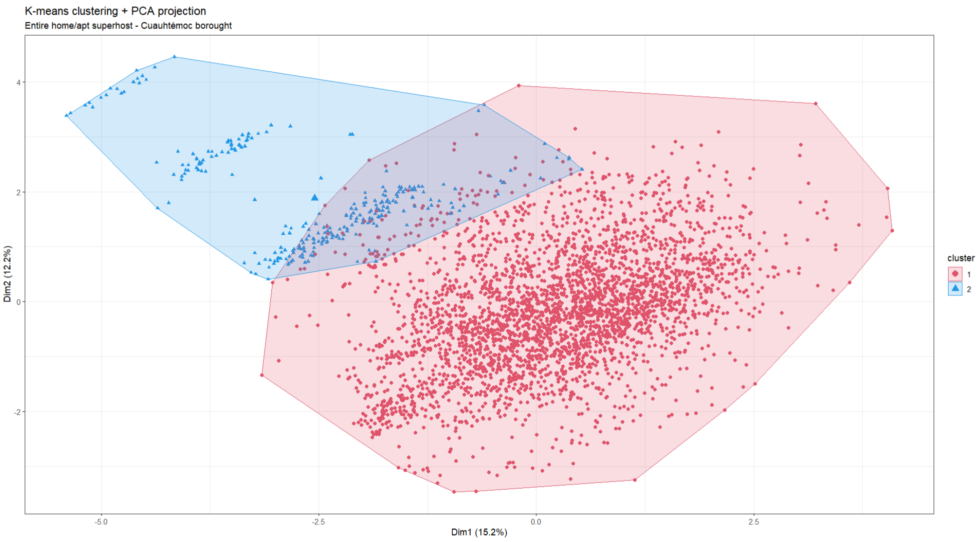

Figure 3.8: Clusters identified for superhosts offering entire home/apt in Cuauhtémoc.

for this arrival period of time is divided into four time slices, corresponding to each quarter of 2023 (excluding the last one) and the first bimester of 2024 (lead time $t \in \{4, 3, 2, 0\}$).

Chapter 4

Model specification

We introduce this section with an informal presentation of the general pricing problem faced by hosts. Consider a seller with one unit of a perishable product, which loses all its value if is not sold by time $t = 0$. The seller starts selling at time $t = T$ and can set a different price in each period. t represents the lead time, which ranges from T to 0; the closer t is to 0, the sooner the expiration date. The seller's goal is to maximize the expected value, defined by the following Bellman equation:

$$V_t = \max_{p_t}(D_t(p_t))p_t + (1 - D_t(p_t))V_{t-1}, \tag{4.1}$$

where V_t represents the expected profit from selling the product at time t. $D_t(p_t)$ is the probability that the item will be sold at price p_t. With probability $1 - D_t(p_t)$ the item remains unsold, and the seller moves into the next period when a new price could be set. $D_t(p_t)$ varies over time, and buyers' willingness to pay for the product may also be time-varying. V_{t-1} is the continuation value and it represents the opportunity cost if the item is sold in period t. If $t = 0$, then there is no continuation value, and V_0 is simply the expected profit from selling the product at price p_0.

Now, what happens if adjusting prices between selling periods is costly for the host? Equation 4.1 can be rewritten as

$$V_t = \max_{p_t}\left\{(D_t(p_t))p_t + (1 - D_t(p_t))V_{t-1} - m_t\mathbb{1}_{\{p_t \neq p_{t+1}\}}\right\}, \tag{4.2}$$

where m_t is the price adjustment cost and $\mathbb{1}_{\{p_t \neq p_{t+1}\}}$ indicates whether the seller decides to change the price from the previous period. When the revenue gained from a price adjustment outweighs its cost, the seller will proceed with the price change. However, if the cost of adjusting the price is too high, the optimal price will remain unchanged despite variations in demand or inventory. Below are the econometric specifications of each object.

4.1 Setup

For a given check-in date, renters are allowed to reserve a listing up to T days in advance. We will only consider whether the listing is rented, or not, for a given arrival date. This assumption preserves the structure of multi-day rentals where a stay of several days can be thought of as multiple individual stays[1].

We will assume that (i) bookings are instant; (ii) there are not cancellations; and (iii) hosts manage their own listings (we will not allow the presence of management services).

4.2 Model of demand

4.2.1 Individual guest's probability of renting a listing

This section characterizes the probability that a customer chooses a specific listing. Renters travel date is assumed to be exogenous. t days before a given arrival date, a renter selects his (her) preferred option among all the available listings by solving a static discrete choice problem. The utility for a renter to choose listing k at time t is

[1] In the model proposed by Pan and Wang (2021), renters are categorized by rental type, depending on the desired check-in day and length of stay for a given check-in week; the total number of rental types consists of 29 different cases.

$$
\begin{aligned}
z_{t,k} &= v_{t,k} + \epsilon_{t,k} \\
&= X_k\beta + \alpha P_{t,k} + \epsilon_{t,k}
\end{aligned} \tag{4.3}
$$

where X_k are the observed attributes of listing k, $P_{t,k}$ is the effective nightly rate of listing k at time t, and $\epsilon_{t,k}$ is the unobserved idiosyncratic utility shock with Type 1 Extreme Value distribution. If the renter exits the market without making a purchase, the utility is normalized to $\epsilon_{t,0}$. Let $G_{t,k}$ be the probability that a customer elects listing k, then $G_{t,k}$ is given by (McFadden (1972))

$$
\begin{aligned}
G_{t,k} &= \mathbb{P}\left(z_{t,k} > \max_{r \neq k}\{z_{t,r}\}\right) \\
&= \frac{\exp\{v_{t,k}\}}{1 + \exp\{v_{t,k}\} + \sum_{r \neq k}^{N_t} \exp\{v_{t,r}\}}
\end{aligned} \tag{4.4}
$$

where N_t is the number of all the available listings at time t.

4.2.2 Customer arrival process

Let us suppose that t days prior to the check-in date customers randomly arrive according to a Poisson distribution with parameter λ_t. Let M_t be the total number of potential renters and $D_{t,k}^{M_t}$ the conditional probability that among all M_t customers, at least one of them is interested in listing k.

To calculate $D_{t,k}^{M_t}$, let $Y_{t,k}$ be a Binomial distribution with parameters $(M_t, G_{t,k})$, thus

$$
\begin{aligned}
D_{t,k}^{M_t} &= \mathbb{P}(Y_{t,k} > 0) = 1 - \mathbb{P}(Y_{t,k} = 0) \\
&= 1 - \left[(1 - G_{t,k})^{M_t}\right]
\end{aligned} \tag{4.5}
$$

Since M_t is not known, it is necessary to compute the unconditional probability of choosing listing k. Then

$$
\begin{aligned}
D_{t,k} &= \sum_{x=0}^{\infty} \mathbb{P}(M_t = x) D_{t,k}^{M_t} = \sum_{x=0}^{\infty} \left(\frac{e^{-\lambda_t}(\lambda_t)^x}{x!} \right) [1 - (1 - G_{t,k})^x] \\
&= \sum_{x=0}^{\infty} \left(\frac{e^{-\lambda_t}(\lambda_t)^x}{x!} \right) - \sum_{x=0}^{\infty} \left(\frac{e^{-\lambda_t}((\lambda_t)(1 - G_{t,k}))^x}{x!} \right) \\
&= 1 - e^{-\lambda_t} \sum_{x=0}^{\infty} \left(\frac{((\lambda_t)(1 - G_{t,k}))^x}{x!} \right) \\
&= 1 - e^{-\lambda_t} \cdot e^{(\lambda_t)(1-G_{t,k})} = 1 - e^{-\lambda_t} \cdot e^{\lambda_t - \lambda_t G_{t,k}} \\
&= 1 - \exp\{-\lambda_t G_{t,k}\} \qquad (4.6)
\end{aligned}
$$

$D_{t,k}$[1] represents the analytical form of the demand function of listing k at time t. $1 - D_{t,k}$ is the non-rental probability.

4.3 Host's price-setting problem

4.3.1 State variables

A host's state variables in period t, (s_t, u_t), could be decomposed into (P_{t+1}, a_{t+1}, u_t). P_{t+1} is the price set by the host at time $t+1$, which may influence the host's pricing decision due to a price adjustment cost. a_{t+1} denotes the inventory level of a listing and it characterizes the availability status of the listing for an arrival date. u_t is the price specific idiosyncratic shock which is assumed to be independent and identically distributed (i.i.d.) with Type 1 Extreme Value distribution and scale parameter σ.

4.3.2 State transition

Let $\nu_t(s_{t-1}, u_{t-1} | s_t, u_t, P_t)$ be the transition probability to depict a host's perception of how the states evolve over time; i.e., how inventory evolves given the decisions

[1]This definition differs slightly from the one proposed by Pan and Wang (2021) since we are assuming that bookings are instant and it is not necessary to consider a processing probability, $\tilde{g}_t$.

made when choosing prices. Assuming that conditional independence is satisfied (Rust (1987)), this probability could be broken up as $\nu_t(s_{t-1}, u_{t-1}|s_t, u_t, P_t) = \omega(u_{t-1})g_t(s_{t-1}|s_t, P_t)$, where ω is the density function of the idiosyncratic shock u_{t-1} and g_t is the transition probability of the observed states s_{t-1} given s_t and P_t. g_t could be further expressed as

$$g_t(s_{t-1}|s_t, P_t) = h_t(a_t|s_t, P_t) \tag{4.7}$$

where $h_t(a_t|s_t, P_t)$ is ruled by the rental probability $D_{t,k}$. Assuming that no cancellations occur once a listing has been reserved, the transition matrix h_t for the availability indicator is given by

		a_t	
		0	1
a_{t+1}	0	1	0
	1	$D_{t,k}(P_t, s_t)$	$1 - D_{t,k}(P_t, s_t)$

At time t, a_t can take the values 1 and 0, indicating whether the listing k is available, or not, respectively. If $a_{t+1} = 0$, the probability that the listing k is still rented at time t is equal to 1, given that the model does not allow cancellations. Conversely, if at time $t + 1$ the listing k is available (i.e., $a_{t+1} = 1$) then the probability that it is rented in the next period would be $D_{t,k}(P_t, s_t)$, which depends on the price chosen at time t. With this,

$$\nu_t(s_{t-1}, u_{t-1}|s_t, u_t, P_t) = \omega(u_{t-1})h_t(a_t|s_t, P_t). \tag{4.8}$$

4.3.3 Bellman equation

Considering $D_{t,k}$, forward-looking hosts aim to maximize their anticipated revenues for each booking date within the series of selling periods. During each subsequent selling period, hosts have the option to either modify their pricing or maintain the previous price; however, adjusting the price incurs a cost. Based on Pan and Wang (2021), we assume that prices are discrete and there are idiosyncratic shocks specific to prices, which are observable only by the hosts. For a given listing k, the per-period payoff function is

$$D_t r_t,$$

where D_t is the probability given by equation 4.6, r_t is the revenue associated to time t. Therefore, for a given listing k and arrival date, the optimal pricing decision is summarized by the following Bellman equation

$$\begin{aligned} V_t(s_t, u_t) = \max_{P_t} \Big\{ & D_t(P_t, s_t) r_t(P_t, s_t) \\ & + (1 - D_t(P_t, s_t))\, \mathbb{E}\left(V_{t-1}(s_{t-1}, u_{t-1}) | s_t, u_t, P_t\right) - m_t \mathbb{1}_{\{P_t \neq P_{t+1}\}} + u_t(P_t) \Big\} \end{aligned} \tag{4.9}$$

In the definition of the state variables, once $a_t = 0$, the value function $V_t(s_t, u_t) = 0$.

4.4 Equilibrium

A host's pricing decision relies on his (her) individual conditions and notions about the evolution of these conditions over time. The host's individual conditions incorporate his (her) inventory status and the price defined in the previous period. Under all these premises, we could define the following equilibrium.

Definition 4.1. *The equilibrium consists of the demand probabilities* $\{D_{t,k}\}$*, the pricing functions* $\{P_{t,k}\}$*, and the initial availability vectors* $\{a_{T+1,k}\}$ *such that*

1. *Given the current period availability* $\{a_{t+1,k}\}$ *and last period price* $\{P_{t+1,k}\}$*, each listing owner* k *chooses his (her) price according to the pricing function* $\{P_{t,k}\}$ *which solve equation 4.9.*

2. *In each period* t*, listings are rented according to the demand probability* $\{D_{t,k}\}$*.* $\{D_{t,k}\}$ *also determine the next period availability vector* $\{a_{t,k}\}$*.*

3. $\{a_{t+1,k}\}$ *determine the number of players in each period and* $\{P_{t,k}\}$ *determine the pricing decision.*

The existence of this equilibrium arises from the finite horizon and finite action-space problem, discussed by Maskin and Tirole (2001).

Chapter 5

Estimation

We are interested in estimating demand and supply parameters, $\theta_1 = \{\beta, \alpha, \{\lambda_t\}\}$ and $\theta_2 = \{\{m_t\}, \sigma\}$, respectively. In order to have a parsimonious model and to implement unconstrained optimization, we imposed the following restrictions on $\{\lambda_t\}$, $\{m_t\}$, and σ parameters:

- Customer arrival parameters $\{\lambda_t\}$ capture increases in demand over time and these could depend not only on the lead time, but also on the arrival date. We reduced this parameter set to consider only two possible values: $\lambda^{\text{ weekday }}$ and $\lambda^{\text{ weekend }}$. Since these parameters must be non-negative, we defined $\lambda^j = \exp\{\delta^j\}$, for $j \in \{\text{ weekday }, \text{ weekend }\}$ and estimate δ^j.
- Menu costs parameters $\{m_t\}$ represent the costs of adjusting prices over time and these depend on lead time. If we have T time slices for a given arrival date, $\{m_t\}$ consists of $\{m_{T-1}, m_{T-2}, ..., m_1, m_0\}$. We also defined $m_t = \exp\{\mu_t\}$ and estimate μ_t.
- The scale parameter for idiosyncratic shock to the host must be non-negative. Similarly, we computed $\sigma = \exp\{\kappa\}$.

With this in mind, we can rewrite the demand and supply parameter sets as $\tilde{\theta_1} = \{\beta, \alpha, \delta^{\text{ weekday }}, \delta^{\text{ weekend }}\}$ and $\tilde{\theta_2} = \{\mu_3, \mu_2, \mu_0, \kappa\}$, respectively. In this specification, β is a vector in $\mathbb{R}^{15}$ corresponding to listing attributes described in table A.1. So, $\tilde{\theta_1}$ comprises a total of 18 parameters, whereas $\tilde{\theta_2}$ contains only 4 parameters.

We estimate the model in two steps, using maximum likelihood estimation and bootstrapped errors. In the initial step, the demand parameters are estimated; next, the supply parameters are subsequently estimated in the dynamic pricing phase. Estimating price adjustment costs is challenging due to the unique nature of each listing, leading to listing-specific value functions. However, the clustering process described in 3.2.1.3 significantly reduced the number of value functions to be employed in the dynamic pricing step.

5.1 Step 1: Demand estimation

Let $a_{t,k}$ be the inventory indicator for listing k at selling period t and $D_{t,k}$ the demand probability given by equation 4.6. Thus, for a given listing k and arrival date, the demand likelihood function is

$$\begin{aligned} L_k^{\text{demand}} &= \mathbb{P}[a_{T,k}, a_{T-1,k}, ..., a_{1,k}, a_{0,k} | a_{T+1,k}; \tilde{\theta}_1] \\ &= \mathbb{P}[a_{T,k} | a_{T+1,k}; \tilde{\theta}_1] \mathbb{P}[a_{T-1,k} | a_{T,k}; \tilde{\theta}_1] \cdots \mathbb{P}[a_{0,k} | a_{1,k}; \tilde{\theta}_1] \\ &= \prod_{t=0}^{T} \left\{ \left(D_{t,k}{}^{(1-a_{t,k})} \right) ((1 - D_{t,k})^{a_{t,k}}) \right\}. \end{aligned}$$

Therefore, for a given listing k and arrival date, the demand log-likelihood function is

$$LL_k^{\text{demand}} = \sum_{t=0}^{T} \{(1 - a_{t,k}) \ln D_{t,k} + a_{t,k} \ln (1 - D_{t,k})\}. \tag{5.1}$$

5.2 Step 2: Pricing

Let

$$v_t^{P_t} = D_t(P_t, s_t) r_t(P_t, s_t) + (1 - D_t(P_t, s_t)) \mathbb{E} \left(V_{t-1}(s_{t-1}) | s_t, P_t \right) - m_t \mathbb{1}_{\{P_t \neq P_{t+1}\}} \tag{5.2}$$

be the specific value function for P_t. If P_t can take H_t different values, then the expected value function in equation 5.2 could be computed as[1]

$$\mathbb{E}(V_t(s_t)) = \sigma \ln \left(\sum_{h=1}^{H_t} \exp\left(\frac{v_t^h}{\sigma} \right) \right) + \sigma\gamma,$$

where γ is the Euler-Mascheroni constant. Since the set of prices H_t is discrete, the choice probability for the h^{th} price is given by (McFadden (1972))

$$\frac{\exp\left(\frac{v_t^h}{\sigma}\right)}{\sum_{r=1}^{H_t} \exp\left(\frac{v_t^r}{\sigma}\right)}.$$

Consequently, for a given listing k and arrival date, the pricing decision log-likelihood function is

$$\begin{aligned} LL_k^{\text{pricing}} &= \ln\left(\mathbb{P}\left[P_{T,k}, P_{T-1,k}, ..., P_{1,k}, P_{0,k} | \{a_{t,k}\}_{t=0}^{T+1}; \tilde{\theta}_2\right]\right) \\ &= \ln\left(\mathbb{P}\left[P_{T,k} | a_{T+1,k}; \tilde{\theta}_2\right] \prod_{t=0}^{T-1} \mathbb{P}\left[P_{t,k} | P_{t+1,k}, a_{t+1,k}; \tilde{\theta}_2\right]\right) \\ &= \ln\left(\prod_{t=0}^{T} \left(\frac{\exp\left(\frac{v_t^{P_{t,k}}}{\sigma}\right)}{\sum_{h=1}^{H_t} \exp\left(\frac{v_t^h}{\sigma}\right)} \right)^{a_{t,k}} \right) \\ &= \sum_{t=0}^{T} \left[a_{t,k} \left(\frac{v_t^{P_{t,k}}}{\sigma} - \ln\left[\sum_{h=1}^{H_t} \exp\left(\frac{v_t^h}{\sigma}\right) \right] \right) \right] \end{aligned} \tag{5.3}$$

As this is a finite horizon problem, backward induction could be used to compute the specific value functions given by 5.2.

5.3 Estimation results

The demand and supply parameters were estimated using the log-likelihoods defined by equations 5.1 and 5.3, respectively, jointly considering the 142 repre-

[1]Straightforward derived from properties listed in appendix A.4.

sentative listings and 60 possible arrival dates. Since the pricing decision problem is host-specific, the set of prices considered by the host in their decision-making must be also host-specific. Therefore, a distinct price grid H_t was defined for each listing. The dimension of this grid includes 10 possible prices, determined by observing actual prices and their variations over time.

The estimation process was implemented from scratch in Python using the maximum likelihood method, optimizing equations 5.1 and 5.3 over the parameter vectors $\tilde{\theta_1}$ and $\tilde{\theta_2}$. For each log-likelihood function, different initial points were randomly chosen to initialize the optimization step and methods *BFGS*, *L-BFGS-B*, *Nelder-Mead*, *Powell*, and *SLSQP* were tested.

5.3.1 Demand parameters

Consistent estimates of demand parameters were achieved using the L-BFGS-B, Powell, and SLSQP methods, all of which also successfully converged in the optimization process. Among these, the SLSQP method demonstrated the best execution times. The estimated parameters are presented in table 5.1. The columns showing the standard deviation, and the minimum and maximum values for each parameter, were obtained using a bootstrap with 3,600 iterations under the SLSQP method. Figures 5.1 and 5.2 display the histograms derived from the bootstrap implementation for the demand parameters $\tilde{\theta_1}$.

The estimated coefficients in table 5.1 generally have the expected signs, with only a few exceptions:

- *neighbourhood*: Instead of favoring locations with the highest concentration of listed properties, the negative sign suggests that customers prefer alternatives away from central areas in Mexico City. This preference could be attributed to the ease of transportation within the city and the rising rental prices in neighbourhoods like Cuauhtémoc. López and Ramírez-Álvarez (2021) noted that access to public transportation routes can lower rental prices by 0.65%; this reduction in price would increase the likelihood of selecting the listing, which is consistent with the sign observed in our estimation.

- *accommodates* and *amenities*: These two coefficients together suggests that people prefer spaces with a small accommodation capacity but sufficient amenities, likely because Mexico City is primarily a destination for business, cultural, and gastronomic tourism.

- *review_scores_rating*: It is counterintuitive for this coefficient to have a negative sign. However, Merino and Muñoz-Rodríguez (2024) pointed out that this result occurs when the endogeneity of prices is not considered in the definition of the discrete choice probability.

In addition, the parameters corresponding to δ_{weekday} and δ_{weekend} imply that the average arrival rate is only slightly higher for weekends. The variations obtained for all parameters are generally very small. While it is uncertain whether the likelihood function is concave, the optimization method seem to converge to the same solution. However, it is possible that we are at a local maximum rather than a global one.

Parameter	**Estimate**	**std**	**min**	**max**
$\beta_{\text{neighbourhood}}$	-0.0205	0.0023	-0.0284	-0.0117
$\beta_{\text{host\_is\_superhost}}$	0.1343	0.0502	-0.0237	0.6065
$\beta_{\text{room\_type}}$	0.3219	0.022	0.2362	0.3964
$\beta_{\text{host\_identity\_verified}}$	-3.5495	0.0434	-3.6889	-3.3871
$\beta_{\text{host\_listings\_count}}$	0.6755	0.0339	0.3538	0.7856
$\beta_{\text{host\_response\_rate}}$	0.1096	0.0188	0.0382	0.1744
$\beta_{\text{host\_since}}$	0.129	0.0097	0.0611	0.1701
$\beta_{\text{host\_verifications}}$	0.0164	0.0253	-0.0707	0.1142
$\beta_{\text{accommodates}}$	-0.2107	0.0197	-0.2893	-0.1262
$\beta_{\text{amenities}}$	4.5484	0.1287	4.1307	4.992
$\beta_{\text{availability\_30}}$	0.3962	0.1317	-0.0824	0.9052
$\beta_{\text{availability\_365}}$	-4.4856	0.14	-5.0178	-3.7553
$\beta_{\text{instant\_bookable}}$	0.1767	0.0416	0.0221	0.3065
$\beta_{\text{number\_of\_reviews}}$	0.041	0.1368	-4.5786	0.3137
$\beta_{\text{review\_scores\_rating}}$	-0.5735	0.0119	-0.6107	-0.4459
$\alpha_{\text{ln\_price}}$	-0.1767	0.0165	-0.2736	-0.1168
δ_{weekday}	4.448	0.0803	4.1573	4.7911
δ_{weekend}	4.5057	0.0812	4.178	4.8671

Table 5.1: Estimated demand parameters $\tilde{\theta}_1$.

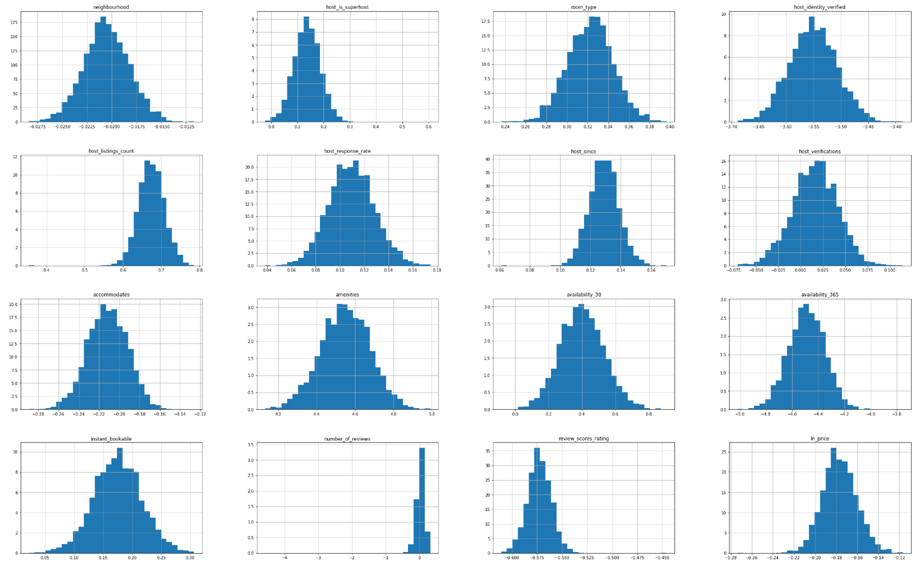

Figure 5.1: Histograms of the demand parameters associated with the listing's attributes.

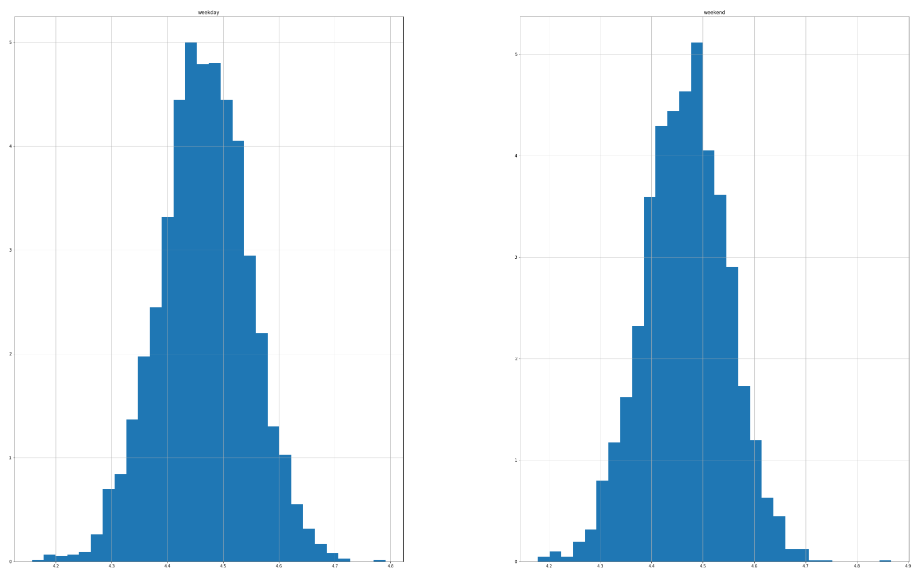

Figure 5.2: Histograms of the demand parameters associated with the arrival customer process.

5.3.2 Pricing decision parameters

After estimating the demand parameters, the supply side parameters were then estimated. Consistent and convergent estimates were obtained using the Nelder-Mead, Powell, and SLSQP methods. The Powell and SLSQP methods yielded the fastest execution times. To maintain consistency with the demand estimation step, the SLSQP method was chosen for the bootstrap implementation. Due to the computational burden of this estimation, derived from the recursive relationship in the value functions defined in log-likelihood 5.3, the number of bootstrap iterations was reduced to 1,200. To prevent Python overflow errors during this step, the value functions were computed in units of thousands (of Mexican pesos, MXN).

The estimated supply parameters $\tilde{\theta}_2$ are presented in table 5.2 and its histograms are displayed in figure 5.3. In table 5.2 it can be observed that the variability of these estimators is significantly greater compared to the demand case. Figure 5.3 shows that the distributions of these estimators generally tend to be left-skewed; however, they exhibit heavy tails on the right. This indicates that the pricing log-likelihood function is highly sensitive to changes in the sample, which may be due to our data failing to adequately capture the dynamics of menu costs. It would be advisable to have data with sufficient frequency to mitigate the effects of censorship in calendar data. This, however, involves a trade-off between information quality and computational cost.

Table 5.3 presents the estimated values under the original parametric specification θ_2. The estimated value for the scale parameter of the logit shock is very similar to the one found by Pan and Wang (2021) in their study. Nevertheless, the authors noted that in Manhattan, the costs of adjusting prices decrease monotonically as the lead time approaches zero. For Mexico City, our findings indicate that this pattern does not occur. Instead, the price adjustment costs remain particularly high throughout the selling period, with a significant decrease only when the lead time is zero. This suggests that hosts are only willing to change their prices when the arrival date is very close, generating substantial price stickiness in this market.

The estimated parameters in table 5.3 are expressed in thousands of MXN. To put these values in perspective, if we assume $\bar{p} = 900$ MXN, which is the average nightly price in our sample, the menu costs exceed 300% of the listing price, except at time $t = 0$, when these costs drop to 87%. As the listing price increases, the percentage represented by the menu costs decreases. The percentages found by Pan and Wang (2021) for Manhattan are between 0.89-2.29% of the average nightly price.

Parameter	**Estimate**	**std**	**min**	**max**
μ_3	0.9907	1.0556	0.0582	8.6908
μ_2	1.1188	0.9460	-0.2423	11.2829
μ_0	-0.2503	1.1575	-0.2503	8.6365
κ	-0.0062	0.7945	-3.7642	5.4457

Table 5.2: Estimated supply parameters $\tilde{\theta}_2$.

Parameter	**Estimate**	**%** w.r.t. $\bar{p}$
m_3	2.6931	299%
m_2	3.0612	340%
m_0	0.7786	87%
σ	0.9938	-

Table 5.3: Estimated original supply parameters θ_2.

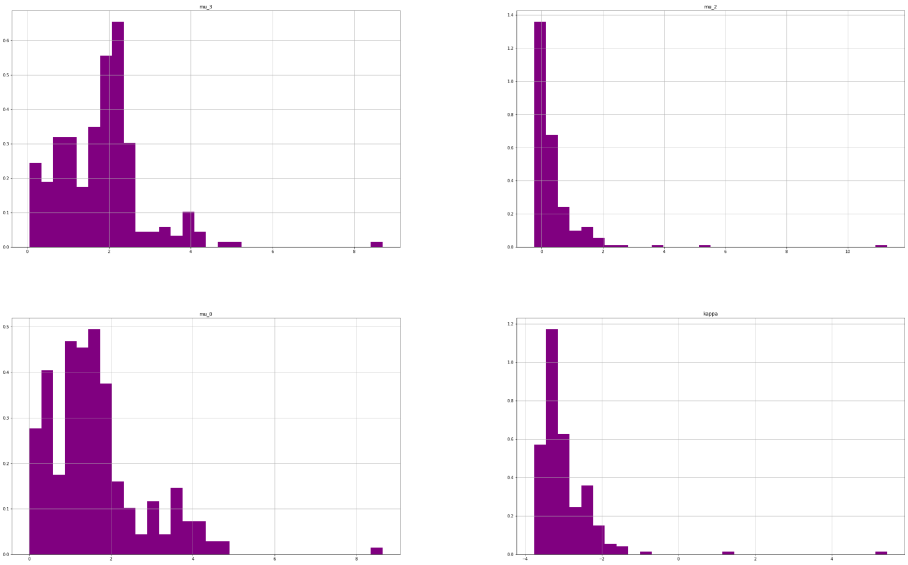

Figure 5.3: Histograms of the supply parameters.

Chapter 6

Conclusions

This research work contributes to the existing literature on Airbnb price dynamics, specifically focusing on Mexico City. While previous studies have examined this market, to the best of our knowledge, this study is the first to incorporate an econometric specification of dynamic pricing that considers the costs hosts may incur when adjusting prices over time. The primary goal of this study was to quantify these menu costs and provide a better understanding of price rigidity in Mexico City.

The estimation process was divided into two steps. The first step involves estimating the demand-side parameters, while the second focuses on the supply-side parameters. The demand-side estimates are consistent and align well with results from previous studies. Additionally, the parameters for the average rate of customer arrivals in Mexico City are intuitive, indicating a slight increase in demand for listings on weekends.

On the supply side, the estimated parameters exhibit greater variance compared to the demand parameters and are more sensitive to changes in the sample used for estimation. This sensitivity may be attributed to the fact that the study relies on only four data points in time to estimate menu costs. Therefore, it is advisable to obtain data with higher frequency, even though this might involve a trade-off with higher computational load.

The estimated menu costs that hosts face when adjusting their prices were expected to decrease monotonically as the arrival date approached. However, the main findings of this research for Mexico City revealed that these adjustment costs are significantly high relative to the average rental price, exceeding 300% of its value. They only decrease substantially when the arrival date is very close, reaching 87% of the average rental price. We can compare these percentages with those found in Manhattan by Pan and Wang (2021), which are less than 3%.

Lower menu costs are generally beneficial for improving overall market efficiency. To achieve this, implementing automated pricing systems and dynamic pricing algorithms can minimize the need for manual intervention by hosts. While these tools are currently available to superhosts on the platform, it would be advantageous to make them accessible to all hosts or to lower the requirements needed to use these tools. Additionally, restructuring the pricing interface (as described in section 3.1) could help reduce price frictions. Simplifying the update process after initial prices are set can make future price adjustments easier and less costly. For instance, using broad pricing categories instead of individual prices for each arrival date can decrease the administrative costs of updating prices.

The guidelines suggested for future research include: (i) estimating menu costs by distinguishing between professional and non-professional hosts, with the expectation of lower costs for professional hosts; (ii) addressing price endogeneity in the demand model specification; (iii) conducting counterfactual studies; (iv) exploring different alternatives in discrete choice models and arrival processes, with common specifications in the literature including nested logit, latent class/mixed logit models, and counting processes such as binomial and negative binomial; (v) considering cancellations over time, similar to the model proposed by Cho and Rust (2018); lastly, (vi) in this study, revenues were estimated directly using the listing price since information on cleaning fees or marginal costs was unavailable; including this information would help profile the price adjustment costs even better.

Appendix A

A.1 Listing attributes description

Table A.1 provides a brief overview of the variables used in estimating the discrete choice model. These variables were selected based on their relevance following the variable clustering analysis.

Field	Type	Calculated	Scale	Description
listing_id	Integer	No		Airbnb's unique identifier for the listing
neighbourhood	Integer	Yes	{1, 2, ..., 16}	Ordinal variable according to table A.2
host_is_superhost	Integer	Yes	{0, 1}	Indicates if the host is a superhost
room_type	Integer	Yes	{1, 3}	1 is for private room; 3 is for entire home/apt
host_identity_verified	Integer	Yes	{0, 1}	Indicates whether the host has a verified identity
host_listings_count	Integer	Yes	{0, 1}	1 if host has more than one listing; 0 if not
host_response_rate	Integer	Yes	{-1, 0, 1, 2}	Original value was categorized; -1 is for NA
host_since	Float	Yes	$\mathbb{R}^+$	Host seniority on Airbnb
host_verifications	Integer	Yes	$\mathbb{Z}^+ \cup \{0\}$	Number of host verifications
accommodates	Integer	No	{1, 2, ..., 16}	The maximum capacity of the listing
amenities	Float	Yes	$\mathbb{R}^+$	Number of amenities divided by 100
availability_30	Float	Yes	$\mathbb{R}^+$	Monthly proportion of availability in the last month
availability_365	Float	Yes	$\mathbb{R}^+$	Annual proportion of availability in the last year
instant_bookable	Integer	Yes	{0, 1}	Indicates whether the host accepts automatic bookings
number_of_reviews	Float	Yes	$\mathbb{R}^+$	Original value divided by 100
review_scores_rating	Float	Yes	[0, 5]	Overall rating given by the guests; 0 is for NA

Table A.1: Description of reduced listing attributes.

A.2 Neighbourhood dictionary

The neighborhoods of Mexico City were organized into an ordinal variable, ordering them in ascending order according to the density of listings within each neighborhood. The values assigned in this classification are included as a numerical variable in the model and are shown in table A.2.

Neighbourhood	**Integer value**
Milpa Alta	1
Tláhuac	2
Xochimilco	3
La Magdalena Contreras	4
Iztapalapa	5
Iztacalco	6
Azcapotzalco	7
Gustavo A. Madero	8
Cuajimalpa de Morelos	9
Venustiano Carranza	10
Tlalpan	11
Álvaro Obregón	12
Coyoacán	13
Benito Juárez	14
Miguel Hidalgo	15
Cuauhtémoc	16

Table A.2: Neighbourhood dictionary.

A.3 Optimal listing clustering

The set of arrival dates was limited to the first two months of 2024. For each of these dates, all listings that were available as of 1Q2023 and that had not submitted any cancellations were filtered. The obtained listings must have information for all dates in the set of arrival dates. After excluding shared and hotel rooms, a subsample comprising 9,912 listings was obtained.

Based on listings data for the first quarter of 2023, this subsample was then grouped based on the combination of neighbourhood, superhost indicator, and room type. Table A.3 presents the number of listings along with the optimal number of clusters identified within each subgroup. The total number of clusters detected for Mexico City was 142.

A.4 Type 1 Extreme Value distribution properties

The following properties were obtained from Muñoz-Rodríguez (2023).

Let $\epsilon_1 \sim \text{Gumbel}(\mu, a)$, where μ and a are the location and scale parameters, respectively.

1. The distribution function is given by

$$F_{\epsilon_1} = e^{-e^{\frac{-(\epsilon-\mu)}{a}}}, \epsilon \in \mathbb{R}.$$

2. $\mathbb{E}(\epsilon_1) = \mu + a\gamma$, where γ is the Euler-Mascheroni constant.

3. If $\alpha > 0$ and $T \in \mathbb{R}$, then $\alpha\epsilon_1 + T \sim \text{Gumbel}(\alpha\mu + T, \frac{a}{\alpha})$.

4. Let ϵ_1 and ϵ_2 be two independent random variables such that $\epsilon_1 \sim \text{Gumbel}(\mu_1, a)$ and $\epsilon_2 \sim \text{Gumbel}(\mu_2, a)$, then

$$\max\{\epsilon_1, \epsilon_2\} \sim \text{Gumbel}\left(a \ln\left(e^{\frac{\mu_1}{a}} + e^{\frac{\mu_2}{a}}\right), a\right).$$

5. Let $\{\epsilon_j\}_{j=1}^{J}$ be J mutually independent random variables such that $\epsilon_j \sim \text{Gumbel}(\mu_j, a)$, then

$$\max\{\epsilon_1, \epsilon_2, ..., \epsilon_J\} \sim \text{Gumbel}\left(a \ln\left(\sum_{j=1}^{J} e^{\frac{\mu_j}{a}}\right), a\right).$$

Neighbourhood	Superhost	Room type	Number of listings	Optimal number of clusters
Álvaro Obregón	No	Private room	167	2
Álvaro Obregón	No	Entire home/apt	183	3
Álvaro Obregón	Yes	Private room	76	2
Álvaro Obregón	Yes	Entire home/apt	79	3
Azcapotzalco	No	Private room	30	2
Azcapotzalco	No	Entire home/apt	30	3
Azcapotzalco	Yes	Private room	8	2
Azcapotzalco	Yes	Entire home/apt	53	2
Benito Juárez	No	Private room	389	2
Benito Juárez	No	Entire home/apt	513	2
Benito Juárez	Yes	Private room	164	3
Benito Juárez	Yes	Entire home/apt	267	2
Coyoacán	No	Private room	275	2
Coyoacán	No	Entire home/apt	203	2
Coyoacán	Yes	Private room	124	2
Coyoacán	Yes	Entire home/apt	118	3
Cuajimalpa de Morelos	No	Private room	60	2
Cuajimalpa de Morelos	No	Entire home/apt	101	2
Cuajimalpa de Morelos	Yes	Private room	2	1
Cuajimalpa de Morelos	Yes	Entire home/apt	96	2
Cuauhtémoc	No	Private room	911	2
Cuauhtémoc	No	Entire home/apt	1454	3
Cuauhtémoc	Yes	Private room	261	2
Cuauhtémoc	Yes	Entire home/apt	1431	2
Gustavo A. Madero	No	Private room	72	2
Gustavo A. Madero	No	Entire home/apt	40	2
Gustavo A. Madero	Yes	Private room	17	2
Gustavo A. Madero	Yes	Entire home/apt	49	2
Iztacalco	No	Private room	21	2
Iztacalco	No	Entire home/apt	28	3
Iztacalco	Yes	Private room	23	2
Iztacalco	Yes	Entire home/apt	22	2
Iztapalapa	No	Private room	72	2
Iztapalapa	No	Entire home/apt	23	2
Iztapalapa	Yes	Private room	11	2
Iztapalapa	Yes	Entire home/apt	24	3
La Magdalena Contreras	No	Private room	25	2
La Magdalena Contreras	No	Entire home/apt	26	2
La Magdalena Contreras	Yes	Private room	5	2
La Magdalena Contreras	Yes	Entire home/apt	10	2
Miguel Hidalgo	No	Private room	353	2
Miguel Hidalgo	No	Entire home/apt	696	3
Miguel Hidalgo	Yes	Private room	131	3
Miguel Hidalgo	Yes	Entire home/apt	564	2
Milpa Alta	No	Private room	4	2
Milpa Alta	No	Entire home/apt	10	3
Milpa Alta	Yes	Entire home/apt	1	1
Tláhuac	No	Private room	7	2
Tláhuac	No	Entire home/apt	13	2
Tláhuac	Yes	Entire home/apt	7	3
Tlalpan	No	Private room	170	3
Tlalpan	No	Entire home/apt	107	2
Tlalpan	Yes	Private room	69	3
Tlalpan	Yes	Entire home/apt	44	3
Venustiano Carranza	No	Private room	82	2
Venustiano Carranza	No	Entire home/apt	60	3
Venustiano Carranza	Yes	Private room	25	3
Venustiano Carranza	Yes	Entire home/apt	49	2
Xochimilco	No	Private room	23	3
Xochimilco	No	Entire home/apt	16	3
Xochimilco	Yes	Private room	7	3
Xochimilco	Yes	Entire home/apt	11	2

Table A.3: Optimal number of clusters for Mexico City.

Bibliography

Abrate, G., Sainaghi, R., and Mauri, A. G. (2022). Dynamic pricing in airbnb: Individual versus professional hosts. *Journal of Business Research*, 141:191–199.

Aguirregabiria, V. and Mira, P. (2010). Dynamic discrete choice structural models: A survey. *Journal of Econometrics*, 156(1):38–67. Structural Models of Optimization Behavior in Labor, Aging, and Health.

Alderighi, M., Nicolini, M., and Piga, C. A. (2015). Combined effects of capacity and time on fares: Insights from the yield management of a low-cost airline. *Review of Economics and Statistics*, 97:900–915.

Arcidiacono, P. and Ellickson, P. B. (2011). Practical methods for estimation of dynamic discrete choice models. *Annu. Rev. Econ.*, 3(1):363–394.

Banco de México (2021). El Mercado de Propiedades Residenciales de Alquiler Temporal en México. Reporte sobre las Economías Regionales Enero-Marzo 2021, pp. 28-33.

Bellman, R. E. (2010). *Dynamic Programming.* Princeton University Press, Princeton.

Bertsekas, D. (2005). *Dynamic Programming and Optimal Control.* Number v. 1 in Athena Scientific optimization and computation series. Athena Scientific.

Cho, S. and Rust, J. (2010). The Flat Rental Puzzle. *The Review of Economic Studies*, 77(2):560–594.

Cho, S. and Rust, J. (2018). Optimal Dynamic Hotel Pricing. 2018 Meeting Papers 179, Society for Economic Dynamics.

de Oca, Á. M. M., Sales-Favà, J., and López-Gay, A. (2018). El auge de airbnb en ciudad de méxico: implicaciones espaciales y sociodemográficas. *Coyuntura Demográfica*, 14:89–99.

Den Boer, A. V. (2015). Dynamic pricing and learning: historical origins, current research, and new directions. *Surveys in operations research and management science*, 20(1):1–18.

Eckstein, Z. and Wolpin, K. I. (1989). The specification and estimation of dynamic stochastic discrete choice models: A survey. *The Journal of Human Resources*, 24(4):562–598.

Escobari, D. (2012). Dynamic pricing, advance sales and aggregate demand learning in airlines. *Wiley-Blackwell: Journal of Industrial Economics*.

Gallego, G. and Van Ryzin, G. (1994). Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management science*, 40(8):999–1020.

Gerardi, K. S. and Shapiro, A. H. (2009). Does competition reduce price dispersion? new evidence from the airline industry. *Journal of Political Economy*, 117(1):1–37.

Huang, Y. (2022). Pricing frictions and platform remedies: the case of airbnb. *Available at SSRN 3767103*.

Lazarev, J. (2011). The welfare effects of intertemporal price discrimination: An empirical analysis of airline pricing in u.s. monopoly market s.

López, D. and Ramírez-Álvarez, A. A. (2021). Análisis de Precios Hedónicos para Airbnb en la CDMX. 2021 Working Papers VII, El Colegio de México.

Maskin, E. and Tirole, J. (2001). Markov perfect equilibrium: I. observable actions. *Journal of Economic Theory*, 100(2):191–219.

McClelland, J. and Rust, J. (2018). Strategic timing of investment over the business cycle: Machine replacement in the us rental industry. *Jahrbücher für Nationalökonomie und Statistik*, 238(3-4):295–351.

McFadden, D. (1972). Conditional logit analysis of qualitative choice behavior.

Merino, J. and Muñoz-Rodríguez, E. (2024). Professional Airbnb Hosts in Mexico City: A First Approximation. 2024 Working Papers I, El Colegio de México.

Merino-Zarco, J. J. (2022). *Estimación de un modelo estructural de demanda y oferta para airbnb en la Ciudad de México*. Trabajos de investigación para obtener el grado de Maestría en Economía.

Merlo, A., Ortalo-Magné, F., and Rust, J. (2015). The home selling problem: Theory and evidence. *International Economic Review*, 56(2):457–484.

Misra, S. and Nair, H. S. (2011). A structural model of sales-force compensation dynamics: Estimation and field implementation. *Quantitative Marketing and Economics*, 9:211–257.

Muñoz-Rodríguez, E. A. (2023). Lecture notes in discrete choice models.

Pan, Q. and Wang, W. (2021). Costly price adjustment and automated pricing: The case of airbnb. *Available at SSRN 4077985*.

Prescott, E. C. (1975). Efficiency of the natural rate. *Journal of Political Economy*, 83(6):1229–1236.

Ruiz-Correa, S., Ruiz-Correa, I., Olmos-Carrillo, C., Rendón-Huerta, F. A., Ramirez-Salazar, B., Nguyen, L. S., and Gatica-Perez, D. (2019). Mi casa es su casa? examining airbnb hospitality exchange practices in a developing economy. *ACM Transactions on Social Computing*, 2(1):1–24.

Rust, J. (1987). Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica*, 55(5):999–1033.

Rust, J. (1994). Chapter 51 structural estimation of markov decision processes. volume 4 of *Handbook of Econometrics*, pages 3081–3143. Elsevier.

Stokey, N., Lucas, R., and Prescott, E. (1989). *Recursive Methods in Economic Dynamics*. Harvard University Press.

Su, X. (2007). Intertemporal pricing with strategic customer behavior. *Management Science*, 53(5):726–741.

Sundaram, R. (1996). *A First Course in Optimization Theory*. Cambridge University Press.

Williams, K. R. (2022). The welfare effects of dynamic pricing: Evidence from airline markets. *Econometrica*, 90(2):831–858.

Ye, P., Qian, J., Chen, J., Wu, C.-h., Zhou, Y., De Mars, S., Yang, F., and Zhang, L. (2018). Customized regression model for airbnb dynamic pricing. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 932–940.

Zhang, W. (2022). Optimal dynamic nonlinear pricing for airline networks. *Available at SSRN 3833013*.

Zhao, W. and Zheng, Y.-S. (2000). Optimal dynamic pricing for perishable assets with nonhomogeneous demand. *Management science*, 46(3):375–388.