Essays in Empirical Industrial Organization and Platforms

by

Aaron P. Kaye

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Business and Economics)
in the University of Michigan
2024

Doctoral Committee:

Associate Professor Ying Fan, Co-chair
Professor Francine Lafontaine, Co-chair
Professor Susan Athey
Assistant Professor Zach Brown

Aaron P. Kaye

apkaye@umich.edu

ORCID iD: 0009-0004-9963-6566

DEDICATION

To my parents, Joyce Illfelder-Kaye and Michael Kaye, and my brother, Joshua Kaye.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my incredible dissertation committee, Ying Fan, Francine Lafontaine, Susan Athey, and Zach Brown, for their invaluable support, guidance, and encouragement throughout this long journey. Each played a pivotal role in my development as a researcher, providing detailed feedback on early research proposals, guiding my research direction, providing technical and writing recommendations, and career advice.

Ying's Industrial Organization class reaffirmed my desire to specialize in IO. I am grateful for her advice on big-picture questions and small technical details, honesty about which research avenues seem most promising, and reassurance through many obstacles in research and grad school. Francine's attention to detail and the structure of her guidance helped me get through each step of the dissertation. I greatly appreciate her ability to challenge my thinking and push the boundaries of my research. Her focus on precise language and communication can make a research project connect with a broader audience. I have learned so much from Susan in roles as a research assistant at Microsoft, visitor at Stanford, lab member, mentee, and co-author. I am grateful for her generosity with time and resources. Our time working together has shaped who I am as a researcher today. Zach's IO class set the technical foundation needed to develop my research. He helped me comb through the technical details of my models and troubleshoot the stubborn objective functions. I also appreciate his willingness to suffer the excruciating technical specifics critical to a paper but best detailed in an appendix.

I'd like to thank my co-authors on my second chapter, Sonal Vats and Michael Luca, and the co-authors on my third chapter, Tom Buchmueller, Will Mandelkorn, and Sarah Miller. I'd also like to thank my other co-authors, especially Ayush Kanodia. I am also thankful to the community of faculty in Economics and Business Economics at Michigan. My research benefited from the many Business Economics brown bag presentations.

I also want to recognize the non-academic staff who make research and teaching possible, including Laura Flak, Ashley Stauffer, Analía Gómez Vidal, Lauren Pulay, and Julie Heintz.

I have been fortunate to meet many amazing colleagues and friends during my time in Michigan. The graduate students in Business Economics include Jae Do Choi, Paul Organ, and Emir Murathanoglu, and my fellow IO students, including Xuan Teng, Carolina Tojal

R. dos Santos, Iris Vrioni, Bruna Guidetti, Russell Morton, Eric Parolin, Candice Wang, and David Van Dijcke. Graduate school would not have been the same without my other close friends, Nate and Faith Mather, Caitlin Hegerty, John Tranfaglia, Erin Markiewitz, Brock Rowberry, Amy Ciardiello, Tyler Radler, Kathrine Richard, Dyanne Vaught, Shwetha Raghuraman, Max Hupperz, Maria Aristizabal-Ramirez, Joaquin Cevallos, Kannappan Sampath, Lea Bart, Nikhil Rao, Chris Hollrah, Emily Horton, Jose Moran, Nadim Balague and many others.

Thank you to the professors at Colby College who motivated me to pursue economics, especially my advisors, Michael Donihue and Simge Tarhan. Thank you to my former colleagues and mentors at Microsoft Research, Boston University, and Summit Consulting, who helped prepare me for graduate school. I am also thankful to the members of the Golub Capital Social Impact Lab.

I would also like to acknowledge Black Diesel, York, and Roos Roast for the years of caffeine used to complete this dissertation.

Finally, I owe a heartfelt thank you to my parents, Joyce and Mike, my brother, Joshua, my girlfriend, Lindsay, extended family, and friends back home for all their love and support.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE

LIST OF TABLES

TABLE

LIST OF APPENDICES

ABSTRACT

This dissertation contains three essays in empirical industrial organization, focusing on online platforms and health. In multisided markets, decisions by one party influence the choices and economic outcomes of others. This dissertation considers three such settings. The first chapter studies the welfare effects of personalized product recommendations in the online hotel market. The second chapter investigates the impact of online ratings in the market for primary care physician services. The third chapter analyzes how short-term government subsidies in insurance markets shape long-term market outcomes.

In many online markets, platforms engage in platform design by choosing product recommendation systems and selectively emphasizing certain product characteristics. In Chapter 1, I analyze the welfare effects of personalized recommendations in the context of the online market for hotel rooms using clickstream data from Expedia Group. This paper highlights a tradeoff between match quality and price competition. Personalized recommendations can improve consumer welfare through the "long-tail effect," where consumers find products that better match their tastes. However, sellers, facing demand from better-matched consumers, may be incentivized to increase prices. To understand the welfare effects of personalized recommendations, I develop a structural model of consumer demand, product recommendation systems, and hotel pricing behavior. The structural model accounts for the fact that prices impact demand directly through consumers' disutility of price and indirectly through positioning by the recommendation system. I find that ignoring seller price adjustments would cause considerable differences in the estimated impact of personalization. Without price adjustments, personalization would increase consumer surplus by 2.3% of total booking revenue (~\$0.9 billion). However, once sellers update prices, personalization would lead to a welfare loss, with consumer surplus decreasing by 5% of booking revenue (~\$2 billion).

In addition to recommendation systems, online platforms often incorporate reputation systems with online ratings, the focus of Chapter 2. Chapter 2 focuses on Zocdoc.com, a unique website that integrates physician profiles, patient reviews, and appointment scheduling onto a single platform. We collected data from the website every day for over a year to construct a novel dataset consisting of profiles, reviews, and ratings for primary care physicians in eight metropolitan divisions. We infer bookings from daily records of appointment

availability. Zocdoc displays ratings on a scale of one to five stars, with overall average ratings rounded to the nearest half-star. We use a regression discontinuity design to identify the causal impact of reviews on patients' choice of physician. Our results suggest that patients care quite a bit about quality. However, due to physicians' capacity constraints and the level of demand, 4, 4.5, and 5-star doctors find most of their offered appointments are booked. The main distinction is timing, with lower rating physicians' appointments booked once the appointments with higher-rated physicians become scarce. We find approximately a doubling in patient volume across the cutoff from 4.5 to 5 stars. We conclude by evaluating the differential impact of ratings, finding that the effects are higher for women physicians and physicians with more reviews. We find a small but insignificant difference for hospital affiliate physicians.

Chapter 3 investigates how government subsidies in insurance markets shape long-term market outcomes. We do this by analyzing the Medicare Advantage (MA) program, a public health insurance program in which the federal government provides private insurance companies subsidies to cover Medicare beneficiaries. While these subsidies originally varied sharply at a population cutoff (see Duggan, Starc, and Vabson 2016), recent changes from the Affordable Care Act greatly reduced this discontinuity. We find that, even ten years after this phase-out began, counties that fell right above the population threshold and received higher subsidies before the Affordable Care Act, continue to experience significantly higher rates of plan enrollment, greater MA penetration in the Medicare market, more participating MA plans, and lower outpatient MA prices when compared to counties that just missed these higher subsidies in the past. Our results demonstrate that subsidies can have impacts on health insurance markets that persist after the subsidies are largely phased out. In addition, we find that these subsidies can impact insurer-provider bargaining and reduce the price of outpatient care, especially where insurance markets are highly concentrated.

CHAPTER 1

The Personalization Paradox: Welfare Effects of Personalized Recommendations in Two-Sided Digital Markets

1.1 Introduction

E-commerce has an increasingly prominent role in the economy and continues to reshape economic activity across industries. Consumers turn to online platforms for retail, travel, groceries, dining, medical care, and other goods and services. In the US, 275 million users shop through e-commerce platforms.[1] As of 2022, e-commerce accounted for 16.1% of total retail sales in the United States and was projected to grow to 21.9% by 2025.[2] In the global travel industry, online booking revenue reached $475 billion in 2022 and is projected to surpass $1 trillion in 2030.[3] A few large platforms often dominate each of these industries, design the marketplace, and connect consumers to third-party sellers. A central activity of these platforms is platform design, whereby platforms select product recommendation systems and selectively emphasize certain product attributes. These platforms leverage detailed personal data to inform product recommendations.[4]

This paper considers the welfare effects of personalized product recommendations in two-sided digital markets where platforms design the marketplace but third-party sellers determine prices. In such markets, the effect of personalized recommendations on consumer welfare, platform profit, and seller outcomes is unclear when we consider how personalization changes seller behavior. Platform design decisions impact both consumer decision-making

[1] https://www.statista.com/statistics/273957/number-of-digital-buyers-in-the-united-states/

[2] https://www.statista.com/statistics/379112/e-commerce-share-of-retail-sales-in-us/

[3] https://www.statista.com/statistics/1179020/online-travel-agent-market-size-worldwide/

[4] Large platforms now provide recommendation systems as a service, such as Microsoft Intelligent Recommendations and Amazon Personalize, making sophisticated recommendation systems possible for platforms of all sizes.

and seller pricing incentives (Dinerstein et al., 2018). Recommendation systems present a tradeoff between match quality and price competition. Consumer welfare depends on the quality of the match between consumer tastes and the products that they ultimately choose, the search costs incurred to find products, and prices. Personalized recommendations may improve consumer welfare through the long-tail effect, whereby consumers match with products that more closely align with their individual tastes and they can find these products with less costly search. However, consumer welfare also depends on prices. Personalized recommendations could recommend products to consumers who are better matched and have higher willingness to pay, possibly incentivizing sellers to increase prices. In sum, under personalized recommendations, consumers could be matched to better products but face higher market prices.[5]

Against this backdrop, I ask: what are the welfare effects of personalized recommendations and other platform design policies when sellers can adjust prices and consumers update beliefs? The current empirical literature focusing on improving recommendations typically holds prices fixed or applies to situations without relevant prices. I build on this literature by explicitly modeling seller pricing decisions. I address this question in the context of the online accommodation industry, using clickstream data on hotel purchases from Expedia Group, a large online travel agency (OTA).

I first present empirical evidence that informs how I construct the structural model. Using data from an experiment conducted by Expedia, I show that position in search results, also called "slot", impacts demand even when recommendations are randomized. I also show that hidden product features are correlated with slots assigned by the default recommendation system. Together, these facts suggest that slots could impact demand through both search costs and rational expectations about the recommendation system. I later use a structural approach and find evidence for both mechanisms. The last empirical fact is that prices influence product positions. This fact is reflected on the seller side of my structural model, in which sellers account for how their choice of price impacts product rankings.

Next, I develop a structural model of demand, platform product recommendations, and hotel pricing behavior. I present the structural model working through each component, starting with demand. I estimate a rich demand model featuring costly consumer search based on Weitzman (1979), where consumers have rational expectations about product recommendations. Previous empirical literature modeling consumer search has often made

[5]It is important to distinguish this phenomenon from traditional price discrimination. For example, in first-degree price discrimination, sellers could individualize prices based on consumer-specific purchase tendencies or willingness to pay. The key tradeoff that I highlight in this paper relates to market prices, not individual prices. Personalized recommendations modify the product selection presented to consumers, which changes the equilibrium market prices faced by all consumers.

simplifying assumptions about what information is revealed from search and consumer beliefs. However, these assumptions may be problematic for understanding the role of personalized recommendations. In my model, product features can be visible or hidden. This avoids issues highlighted by Abaluck et al. (2020), where assuming consumers are aware of hidden product features could bias results. I build on the models of Ursu et al. (2023), and Morozov et al. (2021), where consumers know a portion of the match quality term prior to search and learn the remainder from search, by introducing a data-driven approach, similar to nested logit, that splits the variance components of the match quality term. This model nests the full information demand model and optimal sequential search. Finally, I allow product recommendations to impact demand through search cost and rational expectations. Consumers may accurately believe that the products recommended by the platform have superior hidden features, a mechanism from which the empirical literature commonly abstracts. I allow consumers to have rational expectations and show that this setup fits the data better than a benchmark model where position impacts only search cost. Failing to capture this issue would bias my estimates; if consumers, in part, search for products higher on the page because they tend to have better hidden features, then we would conclude from a model without hidden features that consumers have extremely high search costs.[6]

I estimate the demand model via maximum simulated likelihood. I use the optimal sequential search rules from Weitzman (1979) to construct the joint likelihood of clicking and booking decisions. I use variation in length of stay to separately identify utility parameters from search cost parameters. And I use consumers' repeated decisions (clicks and purchases) to identify heterogeneity in the utility and search cost parameters.

After modeling demand, I present the second component of the structural model, the platform model. Hotels aiming to maximize profits encounter an elasticity of demand influenced not only by consumer preferences but also by the platform's recommendation system. Specifically, a change in a product's price can shift its position in search results. The platform model aims to reverse-engineer Expedia's default recommendation system, capturing the relationship between price and product rankings. To account for the complexity of the default recommendations system, I use a "model of a model" approach from machine learning and cryptography literature.[7]

The third component of the structural model is the supply-side model of hotel pricing behavior, in which hotels set prices based on the time of stay and time of search to maximize expected profits, considering consumer preferences and the platform recommendation system.

[6] I include a more detailed overview of these model features in the demand appendix

[7] This is also called "model extraction". There is extensive literature documenting approaches for reverse engineering black-box algorithms in a number of settings Papernot et al. (2017).

I use the opportunity cost of having a unit available to sell in the next period as the relevant marginal cost for hotels.[8] My supply-side model captures key features of the accommodation industry: at low occupancy, hotels have economies of scale, at high occupancy, hotels face increasing costs and capacity constraints. I use an instrumental variable (IV) approach to address the endogeneity concern of modeling costs as a function of quantity.

Then, with my structural model of demand, platform recommendations, and seller pricing behavior complete, I turn to developing recommendation systems to evaluate in counterfactual simulations. The Expedia data include observations from the (non-personalized) default recommendation system and the randomly ordered experimental data. To understand the welfare effects of personalized recommendation, I develop four increasingly personalized recommendation systems. For each recommendation system, I use an ensemble of 170 LambdaMARTs, a popular machine-learning algorithm for ranking problems, presented in Burges (2010). As with demand estimation, a challenge in training recommendation systems is that slot influences consumer choices but is also highly correlated with product features. I address this issue by using data from an experiment where Expedia randomly assigned slots. The least personalized recommender uses data only on product features. The next recommender includes additional data on the consumer queries. Consumers actively volunteer this information, such as length of stay and whether they are traveling with children. The next recommender incorporates personal data based on the consumer's location, distance to the destination, and time of search. The most personalized recommendation system includes data on consumer's past purchases, such as the average price and star rating of their previous purchases.

Next, I combine the structural model and personalized recommendations to evaluate counterfactuals. In the counterfactual simulations, I solve for the equilibrium induced by each of the four recommendation systems in three distinct phases: first, the platform updates the recommendation system; second, sellers update prices; and third, consumers update their beliefs about the recommendation system. My outcomes of interest are seller profits, quantity sold, platform revenue, and consumer surplus. By evaluating the four recommendation systems, this analysis helps us understand 1) the welfare effects of shifting from the default to personalized recommendation systems, and 2) the welfare implications of escalating levels of personalization.

I find that ignoring seller price adjustments would cause considerable differences in

[8] Betancourt et al. (2022) focus on the airline industry with a dynamic pricing model. In their setup, the estimated marginal cost is the opportunity cost (option value) of having the unit in inventory in the next period. While I use the same interpretation of marginal cost, I do not model it as explicitly because of two practical constraints: First, they compare two competing airlines firms in a market, while I observe over 700 competing firms. Second, their data include capacity information, while I observe quantities.

the estimated impact of personalization. In the counterfactual simulations, without price adjustments, personalization increases consumer surplus by 2.3% of total booking revenue. As a back-of-the-envelope calculation, if we scale this up by Expedia's total gross booking revenue in the same year, 2013, this is a $0.9 billion gain in consumer surplus. This finding is consistent with results from other papers that find gains in consumer surplus from improved recommendations. I find only small effects on quantity, revenue, and profits.

However, once sellers update prices, personalization ultimately leads to a welfare loss. Through higher markups, hotel profits increase by 4.9%, and quantity decreases by 4.5%. Gross booking revenue remains relatively unchanged. In the counterfactual with the most personalized recommendation system, consumer surplus declines by 5% of booking revenue (approximately $2 billion). This amounts to a net welfare loss, as the decrease in consumer surplus is 190% of the increase in hotel profits.

My findings have important policy implications. Recent policies in the EU, such as the General Data Protection Regulation (GDPR), limit platforms' ability to record consumer data. The Digital Markets Act (DMA) focuses on gatekeeper platforms and includes provisions for algorithmic transparency. However, much of the regulatory attention to platforms focuses on self-preferencing, price discrimination, platform fees, and network size. This paper's results highlight an overlooked concern in e-commerce platform research and regulation: Better recommendation systems can reduce competition and ultimately harm consumer welfare. This is important to consider as e-commerce platforms' access to personal data grows and technological improvements allow platforms to deploy increasingly sophisticated recommendation systems.

This paper also has implications for managers. Consider a platform deliberating a tradeoff between profits and product match quality. With prices held fixed, the platform might be at a point where the tradeoff is obvious, where steering consumers to slightly more expensive products is unambiguously good for profits. However, this paper points out that evaluating this tradeoff is not so simple since changes in the recommendation system, in turn, change prices. It might, in fact, be more profitable for the platform to steer consumers to lower-priced goods.

1.1.1 Background Literature

This paper contributes to four strands of the literature. First, it contributes to the sizeable literature on how information frictions impact markets Stigler (1961), Akerlof (1970), and Diamond (1971) and how digitization reshapes economic activity Goldfarb and Tucker (2019). This paper is perhaps most related to Dinerstein et al. (2018), which explicitly considers a tradeoff between platform design and price competition on eBay. However, this paper differs from Dinerstein et al. (2018) on two dimensions: Dinerstein et al. (2018) focus on homogeneous instead of differentiated goods, and their counterfactual policy is a redesign of the display page instead of the recommendation system.

Second, this work contributes to the literature on feature emphasis in online platforms. One focus of this literature is price obfuscation, for example, through drip-pricing and junk fees (Ellison and Ellison, 2009; Blake et al., 2021).[9] The context of this paper is similar, as drip pricing also impacts the accommodation market and is a focus of regulatory attention.[10] More broadly, prices are one product feature that platforms can make more or less costly for consumers to learn. Gardete and Hunter (2018) focus on the automobile market with data from Shift.com and study which features should go on landing pages and which can be moved to product pages. My paper is similar in that I allow consumers to learn about product features through search and to know the correlation between hidden and visible product features. This paper's analysis of consumer search is most related to Abaluck et al. (2020), which presents discrete choice methods for when consumers are not fully informed of product features. Their model enables the researcher to evaluate counterfactuals that change feature emphasis. This paper complements Abaluck et al. (2020) in that I develop a similar demand model of visible and hidden product features that additionally permits consideration of counterfactuals related to feature emphasis and the recommendation system.[11] My paper is different from Abaluck et al. (2020) since I also make use of click data and my counterfactual centers on the recommendation system.

Third, this paper builds on the literature on platform design centered on position effects and recommendation systems. The significance of position effects is well documented, as evidenced by Ursu (2018) and Greminger (2022), who also use Expedia data. These observations are consistent with the common business strategy of auctioning top advertisement slots in search results.[12] Much of the research on the welfare effects of recommendation systems aims to blend demand methods with product recommendations and consider counterfactual utility-based recommendation systems; these include De los Santos and Koulayev (2016), Ursu (2018), Greminger (2022), and Compiani et al. (2021). In contrast to these works, my approach adopts techniques popular in industry and data science to generate ranking algorithms that one might

[9] Drip pricing and junk fees are unavoidable parts of transaction prices that are hidden from consumers early in the search process. For example, a hotel might charge a resort fee that is omitted from the price displayed on landing pages.

[10] https://www.ftc.gov/news-events/news/press-releases/2023/10/ftc-proposes-rule-ban-junk-fees

[11] Both of our papers also allow for product features that do not impact utility but are correlated with hidden features.

[12] Position effects are also important for sponsored search; Athey and Ellison (2011) present a model of bidding for sponsored link positions.

expect to encounter on a platform that personalizes its recommendation system. Specifically, I use randomized data and an ensemble of LambdaMARTs.[13] In that sense, this paper relates to Donnelly et al. (2020), which evaluates personalized rankings using data from an e-commerce platform that randomly personalized recommendations for some consumers and presented nonpersonalized recommendations to others. I also contribute to this literature by accounting for the supply side. Works in this stream hold prices fixed under alternative hypothetical recommendation systems, whereas I endogenize seller pricing decisions. Incorporating the supply side appears to be essential since I find that personalization can improve consumer welfare if I hold prices fixed, consistent with the literature, but allowing prices to change yields a loss in consumer surplus.[14]

The fourth is the emerging literature on self-preferencing, which examines hybrid platforms that both operate the marketplace and compete within it. Notable contributions include Teng (2022), which analyzes the Apple App Store, and Lee and Musolff (2021), which examines Amazon's promoting itself as merchant over competing third-party sellers offering the same good.[15] Further, Lam (2021), Farronato et al. (2023) , and Reimers and Waldfogel (2023) investigate Amazon's practices of prioritizing its products over those of competitors. These papers use various parametric approaches to document the extent of self-preferencing and model recommendation system behavior in their respective settings. Lee and Musolff (2021), Teng (2022) and Lam (2021) then use these estimates to perform supply-side estimation of seller costs. This paper extends this line of research by introducing a method to reverse-engineer Expedia's default recommendations system. Instead of adopting a parametric methodology to model the platform recommendation system, I use a "model-of-model" technique from machine learning. This approach addresses potential concerns about misspecification associated with parametric representation of a sophisticated algorithm.

1.1.2 Outline

The rest of the paper is organized as follows. I cover the institutional background in section 1.2. In section 1.3, I present a stylized illustration of the tradeoff between match quality and price effects. I discuss the data and setting in section 2.3. In section 1.5, I present

[13]Similar approaches were used by the contestants who won the Yahoo! Learning to Rank Challenge and the Personalized Expedia Hotel Searches contest.

[14]Agrawal et al. (2022) and Moehring (2023) also consider personalization and use industry-standard approaches to developing recommendation systems. Prices do not play a role in these papers, as they focus on consumer engagement—Agrawal et al. (2022) in educational technology and Moehring (2023) on r/News on Reddit.

[15]This self-preferencing is implemented through Amazon's Buy Box, which is the primary purchase option on a given product page. Self-preferencing through the Buy Box means that Amazon gives itself an advantage in terms of being selected as merchant over third-party sellers of the same good.

three empirical facts that inform the structural model. Section 1.6 introduces the structural model. Section 1.7 presents the estimation strategy and results for the structural model. In section 1.8, I develop personalized recommendation systems. Section section 1.9 presents the counterfactual simulations and results. Last, section 1.10 provides concluding remarks and discusses the next steps for this project.

1.2 Institutional Background

In digital markets, platforms serving as online intermediaries lead to market behaviors that differ significantly from those in traditional, analog environments. A key factor in this difference is the role of information frictions, which are crucial in shaping consumer demand and market outcomes. In conventional economic models, demand is understood as being influenced by consumer preferences, product characteristics, and information frictions. These frictions affect cumulative demand faced by sellers, as the information available to consumers and the cost of acquiring new information directly impact their purchasing decisions.

E-commerce platforms have notably reduced these information frictions. However, they also uniquely influence these frictions through their platform design strategies. Two primary aspects of such design are recommendation systems and feature emphasis. Recommendation systems influence which products consumers are exposed to and the sequence in which they appear. Meanwhile, feature emphasis affects the visibility of specific product attributes by highlighting them in search results or placing them more discreetly on product-specific pages. This control over information flows allows platforms to act as gatekeepers, a role that has garnered significant policy and regulatory attention.

Platforms have incentives to improve their design since improvements can increase purchase volumes and help them respond to competitive pressure from other platforms. One avenue to improve platform design is to enhance the quality of recommendation systems through personalization. Platforms collect massive amounts of consumer data, including purchase histories and other browsing information, and can use these data to personalize product rankings. Addressing the problem of how best to recommend products is the focus of a growing body of literature, subject of data science competitions, and focus point for platforms. Platforms also face a tradeoff between recommending the products most relevant to consumers and recommending the products most profitable for the platforms themselves.

While this paper focuses on the accommodation industry, it addresses a familiar dynamic between sellers and e-commerce platforms in the increasingly digital economy. The platform chooses its platform design, including the recommendation system, but third-party sellers, in this case hotels, set prices. This setup is common among e-commerce platforms. In the accommodation industry, platforms operated by Expedia Group, Booking Holdings, and Airbnb curate listings by third-party sellers who choose prices. In the food and grocery delivery space, platforms, including Instacart, DoorDash, and Grubhub, choose their design, but restaurants and grocery stores choose prices. StubHub and Ticketmaster act as intermediaries in the market for event tickets, yet third-party sellers choose prices. This dynamic also impacts hybrid platforms such as Amazon, which competes as a seller on its own platform but for which third-party sellers constitute almost 60% of its sales.[16]

The hotel industry is an ideal setting in which to study the welfare effects of platform design for several key reasons. First, the online travel booking and accommodations industry is inherently worth studying given its size and economic significance. For example, Expedia's economic footprint can be seen in its global gross booking revenues, which totaled $107.87 billion in 2019.[17] Second, this industry can provide insight into other major industries because of its parallels with broader e-commerce dynamics. This setting shares key characteristics with other popular e-commerce platforms: a few large platforms dominate the space, there are many differentiated goods in each market, and, as stated above, third-party sellers set prices. The third reason is data availability. It is rare for platforms to release clickstream data with this level of detail to the public. Fourth, these data also include details on an experiment where the product rankings were randomized. These randomized data are ideal for developing recommendation systems to evaluate counterfactuals. Last, this industry setting holds promise for addressing a recurring challenge in the search literature: it can be difficult to separately identify search costs from preferences, especially when slot and product features are collinear. In this setting, consumers arrive searching for stays of different numbers of nights. This introduces variation in the returns to search, which allows me to separately identify preferences from search costs.

1.3 Stylized Example

This section illustrates a simplified version of the trade-off between match quality (long-tail effect) and price competition. Figure 1.3.1 is a stylized illustration focusing on one niche hotel, "The Elvis Hotel", and two consumer types: The first type () represents a typical online-shopper. The other type () has a strong affinity and high willingness to pay for the Elvis-themed goods. The typical online-shopper is more common than the Elvis-type, and both types shop for hotels through an online platform. In this setting, the platform

[16] https://www.aboutamazon.com/news/small-business/celebrating-a-record-breaking-holiday-season-for-amazon-with-cust

[17] In the period of study, 2013, Expedia reported $39.44 billion in gross booking revenue. https://www.statista.com/statistics/269386/gross-bookings-of-expedia/

directs consumers to products based on its recommendation system, consumers make purchase decisions, and the hotel chooses a price. The hotel cannot price discriminate between consumer types, so it chooses one price based on the expected set of consumers that would see the hotel.

Figure 1.3.1: Illustration of Welfare Tradeoff: Match Quality vs Price

Note: For illustration purposes not based on data

The figure compares an environment with default (non-personalized) recommendations, and personalized recommendations. Under the *default recommendations*, the set of consumers steered to the hotel would be similar to the population distribution, with the common type ( ) outnumbering the Elvis-type ( ). Under the default recommendations, the hotel faces demand from a set of consumers mostly comprised of the common-type, and optimally chooses a price of $120/night. Under *personalized recommendations*, the platform can identify each consumer's types with some accuracy, and recommend products accordingly. The figure shows the welfare gain from the long tail effect, as it is matching the Elvis-type consumers to the Elvis hotel. However, the hotel, now facing demand from a set of better matched consumers, would have an incentive to increase prices.

This stylized example abstracts away from many complexities of the hotel market; there is no formal model; it ignores entry and exit, consumer arrival to the platform, and the behavior of other hotels and other platforms. However, this example highlights the tradeoff between match quality and prices. In this simple example, the welfare effects of personalized recommendations are unclear; consumers are better matched to products but face higher market prices. We would need a formal model to conclude if the personalized recommendations resulted in a welfare gain or loss.

1.4 Data and Setting

My primary data source is clickstream data from Expedia Group. These data are publicly available on Kaggle.com and were initially released as part of a data science competition hosted through Kaggle and the International Conference on Data Mining (ICDM 2013) to improve Expedia's recommendation system with personalization. These data are popular among data scientists and an increasingly popular resource for researchers, as it is rare for platforms to publicly release such detailed clickstream data.[18]

The data cover searches from November 1, 2012, to June 30, 2013.[19] The data are at the search-impression level, with one observation corresponding to a consumer–product pair. They include 332,344 queries (consumer searches), with 9,917,530 product queries, covering 173 destination countries, and 136,886 unique properties. For each consumer query (a specific consumer's search), the data include details on up to the first 38 product listings, which products were clicked, and which products were purchased. They include characteristics of each hotel, location attractiveness scores, and information about each consumer's specific query and purchase history (summary statistics about past purchases), hotel availability, and prices on nine other OTAs.

The data are organized around hotel searches and impressions and divided into five categories: search criteria, static hotel characteristics, dynamic hotel characteristics, visitor information, and competing OTA information. For instance, search criteria might include the date and time of the search, destination ID, length of stay, number of adults/children/rooms, etc. Static hotel characteristics cover aspects such as hotel ID, country, star rating, user review score, and historical pricing, while dynamic features include the slot (position), promotion indicators, and headline price, among others.

Another important feature of these data is that they include a details from randomized controlled trial. For two-thirds of the data, consumers received results from Expedia's default recommendation system—so-called natural search results. For the other third, consumers received randomly ordered search results. Injecting this type of occasional experimental randomness into search results to train future versions of recommendation systems is a common practice among platforms. However, the results of these experiments are rarely made publicly available to researchers. For this paper, I use the naturally ordered results to estimate

[18]These data have also been used by Ursu (2018), Abaluck et al. (2020), Greminger (2022), and Reimers and Waldfogel (2023).

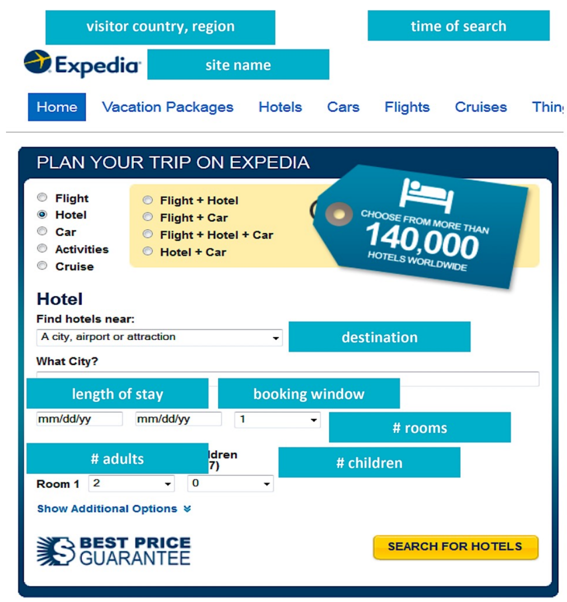
[19]The searches can be for stays as late as October 24, 2014.

demand and the randomly ordered results to train (estimate) alternative recommendation systems that I use in the counterfactual analysis.

1.4.1 Consumer Search Process

Here, I outline the consumer search process and the associated data included and omitted from the Expedia data. We can think of the search process as including three phases: query, search, and purchase. During the query phase, consumers initially input specific search criteria such as location, dates, length of stay, and details about rooms, adults, and children. During the query phase, Expedia also records certain consumer-specific information, such as the country from which the search is made, the booking window (time between the date of search and the date of the stay), the time of the search, and information about the consumer's purchase history.

Figure 1.4.1: Query



Following the query, Expedia displays products on the landing page ordered into slots according to the recommendation system. On the landing page, consumers see the property star rating (class of hotel), customer review scores, approximate location information, whether the property is on sale through a promotion, and the headline price. The headline price is typically the average nightly price of the cheapest available room. Consumers also see a profile picture for each property, but this information is not included in the data.

In the search phase, consumers click on products to navigate to the product page, which reveals more information, including specific product information, room options, and additional

Figure 1.4.2: Landing page, products ordered by recommendation system

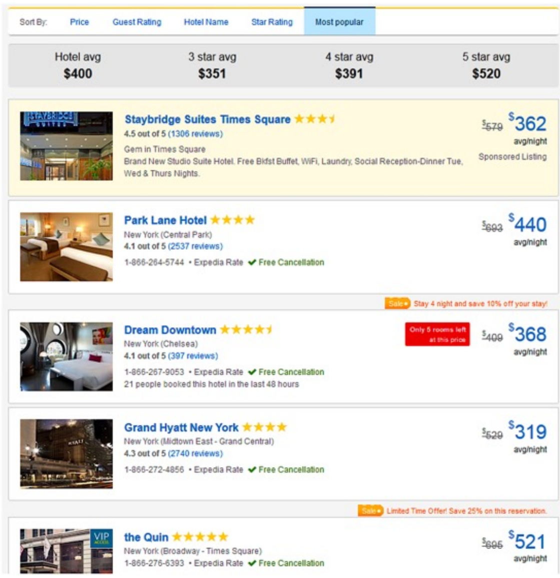
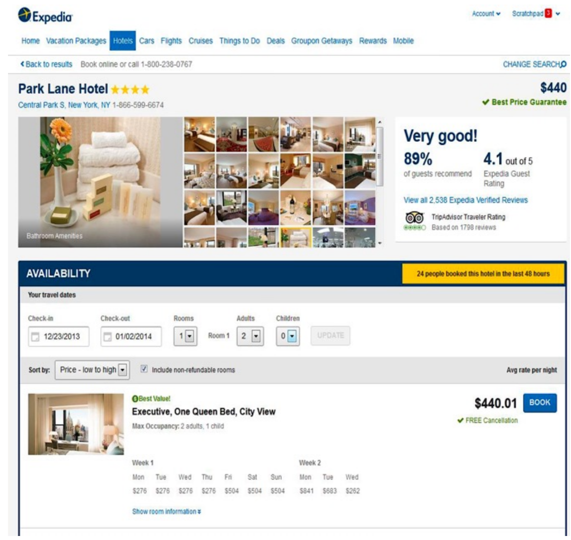


Figure 1.4.3: Click to product-specific page includes hidden product features



pricing details. In most markets, the headline price on the landing page is the nightly rate of the cheapest available room. The Expedia data include two location desirability scores, which capture some of the hotel-specific information that consumers learn through clicks, as the landing page contains approximate but not specific locations.

Finally, in the purchase phase, consumers purchase one of the clicked hotels or end their search (choose the outside option). At this point, Expedia records the gross booking revenue. The final transaction price can be higher than the headline price, as it includes taxes, fees, and upgrades. The differences between the headline and final transaction prices introduce some uncertainty about transaction prices.

1.4.2 Data Processing

Preparing the Expedia data for analysis requires several data processing steps. The Expedia data were released for a data science competition and are well-suited for training recommendation systems. However, a few data limitations present difficulties in conducting the type of demand estimation and counterfactual analysis in this paper. I include additional data processing details in A.1.2).

Market Definitions via K-means clustering. I define markets by groups of search terms. The Expedia data are de-identified, meaning I have hotel and search term identifiers but no keys.[20] I use k-means clustering to group together search terms by the similarity of their search results (details in Appendix A.1.2.1).

Final Transaction Price Prediction. A limitation of this dataset is that it records final transaction prices only when there is a purchase. When a result for a hotel is clicked but no purchase is made, the consumer may still discern the final transaction price, but this price is omitted from the data. Transaction prices are important for two reasons. First, they influence consumer search and purchase decisions. A consumer might, through a click, learn the final price, which informs their next search or purchase decisions. Second, for an accurate measure of consumer welfare, the final prices are essential, as these represent the actual expenditures by consumers. To address this missing data issue, I impute the percent difference between the headline price and the final per-night transaction prices using the hotel-length of stay median (details in Appendix A.1.2.2).[21] Figure 1.4.4, displays the impression level distribution of imputed hidden price difference. In the top market (by revenue), we see a median price difference of 18%, with a thick right tail. We see a bimodal distribution in the second-ranked market with mass points around 0% and 20%. In both cases, we see variation in the pattern of hidden prices both within and across markets.

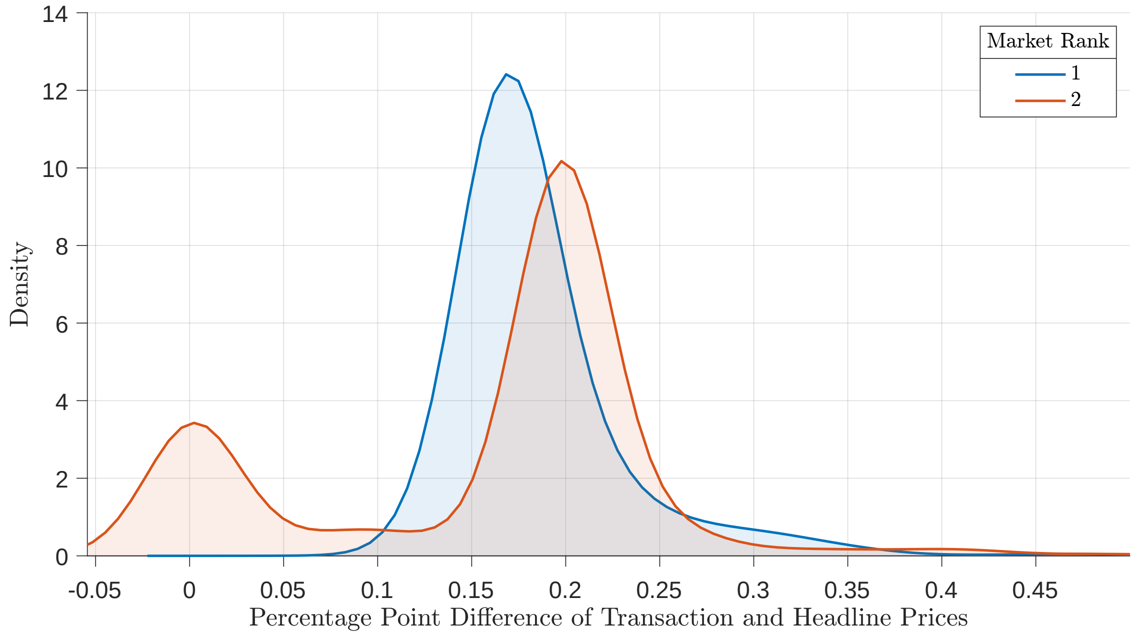
Click order prediction. The data includes indicators for clicks and purchases but does not include information about click orders. Less than 8% of consumer-queries have more than one click. I use a linear prediction model, detailed in Appendix A.1.2.3, to predict the click order.

Sample Selection. Three issues arise from the competition's data sampling method: 1) selection on clicks, 2) oversampling of transactions, and 3) ambiguity in the sample size. I address selection on clicks by using conditional likelihoods in demand estimation and selection weights on the supply side. I address the oversampling of transactions by using sample weights based on reported conversion rates in previous studies. I address the sample size ambiguity by comparing the cross-booking revenue from my data to publicly reported gross booking revenue from the same year. I detail each of the selection issues and solutions in Appendix A.1.3.

[20] For example, while an identifier might indicate "search term 52," there is no direct link to a specific term such as "Manhattan, NY."

[21] I use the hotel median for hotels with a limited number of transactions, while hotels with fewer than 3 observations are assigned the market-length of stay median. A potential concern with this methodology is that hotels could have modified their concealed pricing strategy during the study. I address this in Appendix A.1.2.2.

Figure 1.4.4: Impression Level Kernel Density of Hidden Price Differences by Market



1.4.3 Sample Restrictions

For the primary analysis, I focus on the largest market in terms of revenue. However, I develop the platform model and the recommendation systems I use in counterfactual using data from all five markets. In each of these machine learning applications, jointly estimating the model for each is helpful as since there could be cross-market learning spillovers. In the project's next phase, I intend to expand the analysis to additional markets.

1.5 Empirical Facts, Position Effects, Incentives

As in brick-and-mortar stores, where product placement on shelves (e.g., at eye level) influences consumer decision making, the positioning of products on digital "shelves" in slots on search result pages can influence consumer behavior and seller outcomes. The term "position effects" refers to the influence that position has on consumer behavior and seller outcomes, which is well established in the empirical literature (Ursu, 2018; Greminger, 2022; Donnelly et al., 2020), but its underlying mechanisms are still unclear.

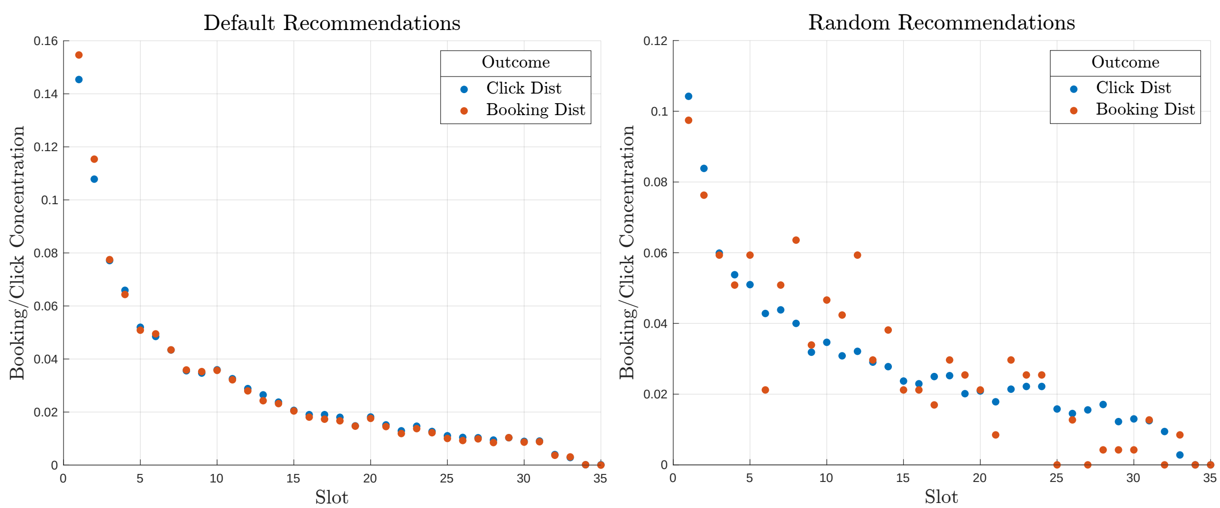
1.5.1 Three Empirical Facts

This section presents three empirical facts that inform how I construct the structural model.

Fact 1: Position impacts demand even when recommendations are random

The first fact documents position effects under default and random recommendations. Figure 1.5.1 plots click and purchase concentrations by slot under Expedia's default recommendation system (left) and under randomized recommendations (right). Blue points denote the percent of all bookings (under the given recommendation system) attributed to the given slot. Similarly, clicks denote the percent of all clicks in the given slot. The figures use concentration instead of levels to avoid misleading comparisons across recommendation systems, as observations with purchases were sampled at different rates for the default versus random recommendations.

Figure 1.5.1: Click and Booking Concentration by Slot and Recommendation System



If consumers were fully informed about products, we would expect to see position effects under the default recommendations (left) but not random recommendations (right). We might still expect to see position effects under default recommendations only due to the correlation between slot and desirable product features. However, since features and slots are uncorrelated in the random rankings, we would expect to see uniformly distributed clicks and purchases. Instead, the data show that clicks and purchases are concentrated in the slots higher on the page, implying that position effects depend on more than visible features. For example, search cost could depend on the slot. Consumers might also have beliefs (rational expectations) about the relationship between hidden product features and slots.

Fact 2: Hidden product features are correlated with slots

The second fact focuses on the relationship between hidden product features and product recommendations. The data include two location desirability scores. These scores can be considered hidden features since a general property location appears on the landing page, but the specific location appears on the product-specific page. Imagine, for example, a consumer searching for a beach vacation. They will see on the landing page that a property is near the beach but can only learn if it is a beachfront property after clicking on the property-specific page. Figure 1.5.2 plots relative location desirability scores by slot. The scores are demeaned on the consumer-query level since location scores can differ from market to market.

Figure 1.5.2: Hidden Features By Slot: Demeaned Location Desirability Scores

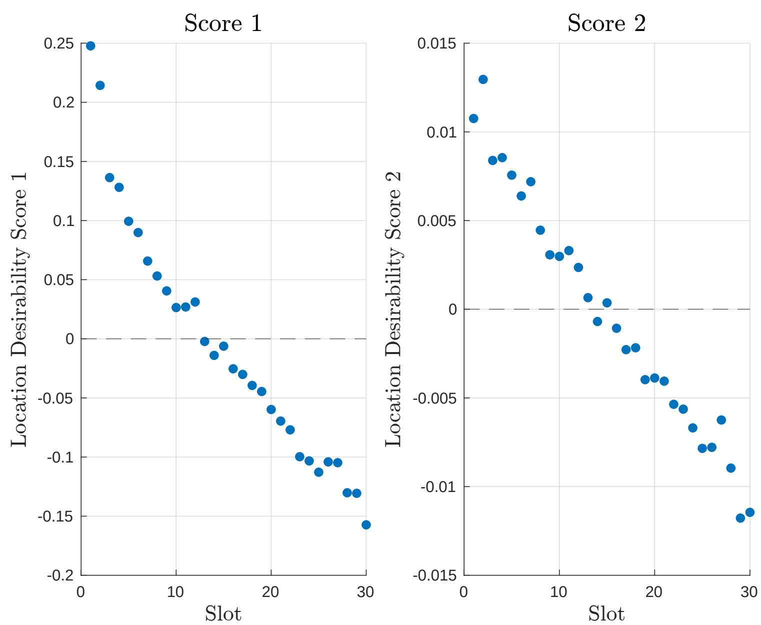


Figure 1.7.1 shows a correlation between slots and location desirability. This means that, on average, a product slotted higher on the page has higher location desirability scores than lower-ranked products in the same search. This correlation is unsurprising since the default recommendation system is likely a function of past consumer decisions, which partly depend on location desirability.

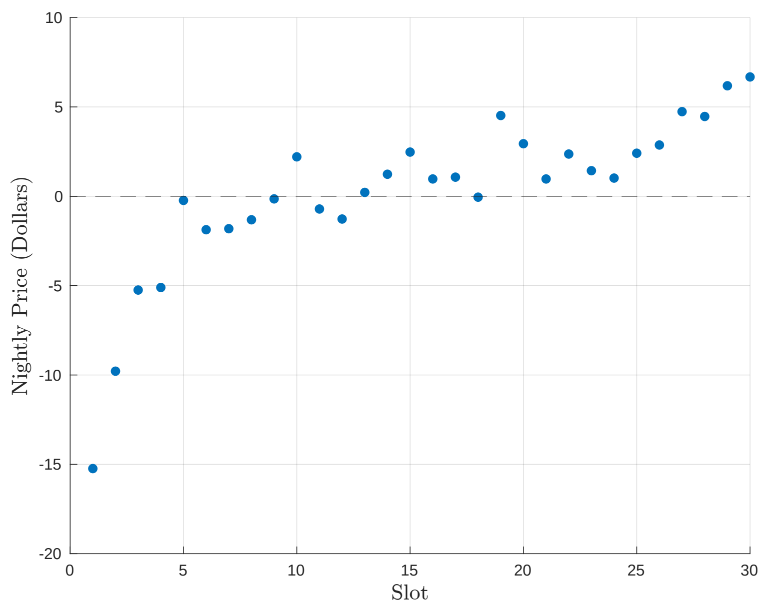
This correlation opens the possibility for a mechanism where consumers "trust the algorithm." In other words, consumers could have the accurate belief (rational expectations) that products positioned higher on the page have superior hidden features. A common approach in the empirical literature assumes that slot only impacts demand through search cost, abstracting from rational expectations about the recommendation system.

It is important to consider the rational expectations mechanism for two reasons. First, to separately identify search cost and utility parameters, practitioners often rely on experimental data with randomly ordered slots. However, with rational expectations, this could be an issue since, by design, consumers do not know they are receiving random recommendations and would still behave based on their beliefs about the default recommendation system. Second, failing to capture beliefs about the recommendation system could bias estimates of search cost; if consumers, in part, search for products higher on the page because they tend to have better-hidden features, then we would conclude from a model without rational expectations that consumers have extremely high search costs.

Fact 3: Price is correlated with slot

This fact centers on the relationship between price and position on the page. Figure 1.7.1 plots the mean relative headline price difference by slot. For example, a value of -\$15 for the first slot implies that, on average, for a given consumer query, the hotel in the first position has a headline price that is \$15 cheaper per night than the average hotel in the search results.

Figure 1.5.3: Headline Price by Slot (Demeaned)



The pattern in these data suggests not only that products in higher slots tend to be lower-priced but also could indicate that Expedia's default recommendation system assigns slots as a function of price—a fact confirmed through my analysis with the platform model. This implies that sellers can potentially improve their slot positioning by lowering prices, and that counterfactual policies altering recommendations could consequently affect prices as well. In light of the apparent relationship between prices and slots, I develop the supply side of the structural model to account for the fact that prices impact demand both directly, through consumers' disutility of price, and indirectly, through positioning by the recommendation system.

1.5.2 Implication of Empirical Facts

The first two facts inform the structure of my demand model and my choice of data. In the primary specification of my demand model, I allow slots to impact consumer decision-making through search costs and rational expectations. I also test these structural assumptions by benchmarking the demand model against a competing one, with the more conventional assumption that slot impacts demand only through search cost.

As for the choice of data, I could use the naturally or the randomly ordered data. For my primary demand specification. I estimate demand using the data from Expedia's default rankings instead of the random rankings. This is for two reasons. First, I can model rational expectations where the beliefs match the data. Second is the added benefit discussed in section A.1.3 of being able to use sample weights informed from other sources.

The third fact informs the supply-side of the model and counterfactual simulations. Since price impacts position, sellers current pricing strategy depends on the default recommendation systems. A change in recommendation systems would, changes the relationship between prices and recommendations, which changes pricing incentives.

1.6 Structural Model

To understand the welfare effects of personalized recommendations, I develop a structural model of consumer demand, product recommendation systems, and hotel pricing behavior. The demand side consists of an optimal sequential search model where consumers have beliefs about the joint distribution of product features and recommendations, form consideration sets through clicks, and make a final purchase decision from their consideration set. For the product recommendation model, I use a "model-of-a-model" machine-learning approach to reverse-engineer Expedia's default recommendation system. Combining the results from the demand and recommendation system models allows me to construct the supply side of the model, where capacity-constrained hotels consider how changes in price impact their position on the page in search results.

1.6.1 Demand

In this section, I describe the individual demand model, an optimal sequential search model based on Weitzman (1979), where consumers have beliefs about the joint distribution of product features and recommendations, form consideration sets through clicks, and make a final purchase decision from their consideration set. The search model has three consumer–product-specific components: indirect utility, search cost, and reservation utility. I use these to construct the final utility and the likelihoods required for estimation. This model requires three basic subcomponents: utility, search cost, and reservation utility. The utility and search cost estimates follow directly from the model parameters and observable, while the reservation utility can be expressed using a value function.

1.6.1.1 Demand Timing

Consumers arrive to the platform exogenously with queries for a specific market and length of stay. The demand model captures consumer search and purchase decisions. Note that I model consumer behavior once consumers are on the platform (including their choice of the outside options), not the decision to search in the first place.[22] Consumers can click, make a purchase, or choose the outside option. The outside option is to end the search without making a purchase.

1.6.1.2 Final Utility

Consider consumer i conducting a search at time t, for a stay at time t', with a length of stay x_{it}^{nights}.[23][24] We can express the consumer's final utility from their search and purchase decisions as:

$$U_{it} = x_{it}^{\text{nights}} u_{ijt}^{\text{choice}} - \sum_{j \in S_{it}} c_{ijt} \tag{1.6.1}$$

where x_{it}^{nights} is the number of nights, u_{ijt}^{choice} is the per-night utility of consumer i's choice, and I subtract the incurred search cost of each product that consumer i added to their consideration set. While utility and reservation utility depend on the length of stay, the search costs that consumers face do not. I will return to this fact in estimation, as this difference allows me to separately identify utility and search costs parameters.

[22]Hortaçsu et al. (2021) develops a demand estimation approach that incorporates consumer arrival and applies it to air travel demand.

[23]The model endogenizes search and purchase decisions but takes consumer arrivals as exogenous.

[24]In most cases, I dropped the t' subscript for clarity, as each consumer search identifies t'.

1.6.1.3 Benchmark Indirect Utility

As a benchmark, it is helpful to consider the full-information demand setup and then highlight how the search model differs. In the full-information demand setup, I assume that the consumer i's utility for product j has two additively separable components:

$$u_{ij} = \delta_{ij} + \varepsilon_{ij} \tag{1.6.2}$$

where δ_{ij} denotes the part of utility observed by the researcher, and ϵ_{ij} represents the portion of utility known to consumers but not observed by the researcher. In full-information demand models, consumers know ϵ_{ij} for each product. In the typical search models used in empirical work, consumers know δ_{ij} and pay a search cost to learn ϵ_{ij}. A few papers have consumers know part of ϵ_{ij} prior to search, and learn part of ϵ_{ij} after search.

1.6.1.4 Indirect Utility Visible and Hidden Product Features

In the context of e-commerce platforms such as Expedia, the assumption that consumers know product features prior to search can be overly strict. Platforms choose their feature emphasis, which determines which product features appear on the landing page and which appear on product-specific pages. Incorrectly assuming that consumers are perfectly informed about product features would bias parameter estimates Abaluck et al. (2020). For example, assuming that consumers are perfectly informed about prices could lead to underestimates of price parameters, as consumers would appear to not react to price differences among unsearched products.

This paper presents a formalized decomposition of indirect utility, distinguishing between "visible" and "hidden" components. Reparametrizing the indirect utility function, we have

$$u_{ij} = \underbrace{\delta^v_{ij} + \varepsilon^v_{ij}}_{\text{Visible}} + \underbrace{\delta^h_{ijt} + \varepsilon^h_{ij}}_{\text{Hidden}} \tag{1.6.3}$$

where δ^v_{ijt} denotes the part of utility observed by the researcher and known to the consumer before search. δ^h_{ijt} is the part of utility observed by the researcher and known to the consumer only after search. Similarly, ϵ^v_{ij} represents the portion of utility known to the consumer prior to search but not observed by the researcher. ϵ^h_{ij} is the portion of utility not observed by the researcher and known to the consumer only after costly search. I detail the structure of the two match quality terms below.

1.6.1.5 Visible and Hidden Variance Components of the Match Quality Term

In the demand model, I distinguish between visible and hidden product features. The match quality term follows a similar structure, with visible and hidden components, but with an added parameter λ that determines how much of the match quality term is known before search and how much is learned from search along with the hidden product features. We can express the sum of the terms as

$$\epsilon_{ijt} = \lambda\varepsilon^{v}_{ijt} + \varepsilon^{h}_{ijt}(\lambda) \tag{1.6.4}$$

where ϵ_{ijt} is consumer i's match quality for product j at time t and follows an i.i.d. type-1 extreme value distribution. ε^{v}_{ijt} is the match quality known before search and follows an i.i.d. type-1 extreme value distribution and is multiplied by $\lambda \in (0, 1)$. ε^{h}_{ijt} follows a Cardell(λ) distribution , whose characteristic function depends on λ.

To achieve this structure, I use a novel application of the properties of the variance components of the type-1 extreme value distribution established in Cardell (1997).[25]. I also use recent advances by Galichon (2022), which proves a relationship between the Cardell distribution and stable distribution. For more details, see Appendix A.3.1.

1.6.1.6 Indirect Utility

Rewriting nightly utility to include the lambda terms, we have the expressions

$$u_{ijt} = \delta^{v}_{ijt} + \delta^{h}_{ijt} + \lambda\varepsilon^{v}_{ijt} + \varepsilon^{h}_{ijt}(\lambda)$$

The value of the outside option is

$$u_{i0t} = \varepsilon_{i0t} \tag{1.6.5}$$

Section 1.7.1.1 details the primary specification of δ^{v}_{ijt} and δ^{h}_{ijt}. Relating this to aspects of platform design, the observable product features that appear on the landing page enter utility through δ^{v}_{ij}, and the product features relegated to the product pages belong to δ^{h}_{ijt}. Similarly, consumers may intuit a portion of the match quality, for example, from product photos or prior searches, which enter through ε^{v}_{ij}. The hidden portion of the match quality term is ε^{h}_{ij}. Consumers know the utility of the outside option, u_{i0t}, prior to search.

[25]These properties are often used to construct nested logit models, such as Berry (1994), which includes nest-level and the item-level variance components of the error term.

1.6.1.7 Search Cost

I assume that the consumer is not fully informed about the hidden components of utility, δ^h_{ijt} and ε^h_{ij}, and must pay a search cost c_{ijt} to learn them.

$$c_{ijt} = f(\theta_i, \text{slot}^{\text{appear}}_{ijt}) \tag{1.6.6}$$

where c_{ijt} is consumer i's search cost for product j. θ_i is the set of consumer-specific search cost parameters, and $\text{slot}^{\text{appear}}_{ijt}$ denotes the position on the page for product j in search it. Advertisements for opaque offers, on occasion, displace products in the slot. Section 1.7 details the functional form of search costs, which allows for heterogeneous search costs and flexibly captures the relationship between c_{ijt} and $\text{slot}^{\text{appear}}_{ijt}$.

1.6.1.8 Reservation Utility in Optimal Sequential Search

Suppose consumer i has already clicked on $r-1$ products. Their maximum utility among the clicked options is $u^*_{i(r-1)} = \max^{r-1}_{k=0}\{u_{ik}\}$. To save on notation, I drop the $r-1$ subscript and let u^*_i refer to the maximum utility among the clicked options at any stage in the search process. The consumer's expected marginal benefit from searching for (in this case, clicking on) item r is given by Weitzman (1979) as:

$$B_{ir}(u^*_i) = \int_{u^*_i}^{\infty} (u_{ir} - u^*_i) f_{u_{ir}}(u_{ir}) du_{ir} \tag{1.6.7}$$

where $f_{u_{ir}}(u_{ir})$ is the probability density of u_{ir}. In the general case, search continues as long as there is a unsearched product where expected benefit exceeds the cost.

$$B_{ir}(u^*_i) > c_{ij} \tag{1.6.8}$$

For each target product, there is an indifference point where $B_{ir}(u^*_i) = c_{ij}$ such that the consumer is indifferent between receiving r_{ijt} with certainty and continuing to search. Under optimal sequential search, consumers search in order of reservation utility. We can define reservation utility, r_{ijt}, for consumer i, product j in search t as the value that satisfies the following equality:

$$c_{ijt} = \int_{r_{ijt}}^{\infty} (u_{ijt} - r_{ijt}) f_{u_{ijt}}(u_{ijt}|\Omega_{it}) du_{ijt} \tag{1.6.9}$$

where r_{ijt} is the level of per-night utility that would make consumer i indifferent between receiving r_{ijt} with certainty or paying search cost c_{ijt} to learn u_{ijt} given information set Ω_{it}. Since consumers know the visible part of utility and learn both δ^h_{ijt} and $\varepsilon^h_{ijt}(\lambda)$, the reservation utility depends on the consumer's beliefs about and the distribution of $\delta^h_{ijt} + \varepsilon^h_{ijt}(\lambda)$. We can rewrite this condition as

$$c_{ijt} = \delta^v_{ijt} + \lambda\varepsilon^v_{ijt} + \int_{r_{ijt}}^{\infty} (\delta^h_{ijt} + \varepsilon^h_{ijt}(\lambda) - r_{ijt} - \delta^v_{ijt} - \lambda\varepsilon^v_{ijt}) f_{u_{ijt}}(\delta^h_{ijt} + \varepsilon^h_{ijt}(\lambda)|\Omega_{it}) d(\delta^h_{ijt} + \varepsilon^h_{ijt}(\lambda)) \tag{1.6.10}$$

A common approach follows Kim et al. (2010), where each utility component except the error term is assumed to be known before search, and search reveals the match quality term. In that setup, $f(u_{ir}|x_{ij})$ depends only on the distribution of the match quality term. Alternatively, in my model, since consumers also learn δ^h_{ijt}, which depends on product features, $f(u_{ir}|x_{ij})$ also depends on the distribution of δ^h_{ijt}.

With some additional algebra, I can rewrite this equation to express reservation utility as

$$r_{ijt} = \delta^v_{ijt} + \lambda\varepsilon^v_{ijt} + E_i[\delta^h_{ijt}|\Omega_{it}] + \zeta_{ijt} \tag{1.6.11}$$

where δ^v_{ijt} and $\lambda\varepsilon^v_{ijt}$ are the visible portion of utility, $E_i[\delta^h_{ijt}|\Omega_{it}]$ is consumer i's expectation of δ^h_{ijt} conditional on their information set Ω_{it}, and ζ_{ijt} is the portion of the reservation utility that does not have a closed-form expression that allows r_{ijt} to satisfy the equality in equation (1.6.10). We can write ζ_{ijt} as a value function

$$\zeta_{ijt} = V(c_{ijt}, \theta^h_i, \lambda, x^{\text{nights}}_{it}|\Phi_{it}, \Omega_{it}) \tag{1.6.12}$$

where the state variables are c_{ijt}, the search cost, θ^h_i, consumer i's utility parameters for hidden product features, λ, and x^{nights}_{it}, the length of stay. Φ_{it} is the distribution hidden utility, and Ω_{it} is the consumer's information set.

In estimation, I solve r_{ijt} numerically since r_{ijt} does not have a closed-form expression.

1.6.1.9 Structural Assumptions

The model requires some structural assumptions, which I document here. As stated earlier, the researcher decides which product features are hidden and which are visible. Additionally, since I treat some product features as hidden, I need structural assumptions about consumers' beliefs about hidden product features. I assume that consumers have rational expectations of hidden utility, with two related components that impact reservation utilities, $E_i[\delta^h_{ijt}|\Omega_{it}]$ and ζ_{ijt}.

For $E_i[\delta^h_{ijt}|\Omega_{it}]$, I assume that consumers form rational expectations of δ^h_{ijt}, conditional on their information set. For the non-price components of δ^h_{ijt} consumers' information set includes the star rating, $slot^{\text{rank}}_{ijt}$, and if the product is on promotion. For the price component of δ^h_{ijt}, the final transaction price, consumers know the headline price and the median percent difference between the headline and final prices.[26] For the ζ_{ijt} component of utility, I assume that consumers know the distribution of $\delta^h_{ijt} + \varepsilon^h_{ijt}(\lambda) - E_i[\delta^h_{ijt}|\Omega_{it}]$.

1.6.1.10 Mechanisms for Position Effects

Now that we have expressions for utility, search cost, and reservation utility, we can discuss how product positioning in the search results impacts demand. The standard approach in the empirical literature imposes a structural assumption that position impacts demand only through search cost. The demand model allows for three mechanisms:

Mechanism 1 Search cost c_{ijt}: Position on the page impacts search cost. This is captured by including $slot^{\text{appear}}_{ijt}$ in the search cost function.

Mechanism 2 Expectation of δ^h_{ijt}: Consumers have accurate beliefs (rational expectations) about the relationship between position and mean hidden utility. This can be captured through $E_i[\delta^h_{ijt}|\Omega_{it}]$ by including the slot in Ω_{it}.

Mechanism 3 Higher-order beliefs: Consumers have beliefs about the relationship between position and the distribution of hidden utility. This can be captured by including the position as an additional state variable in the value function of ζ_{ijt}.

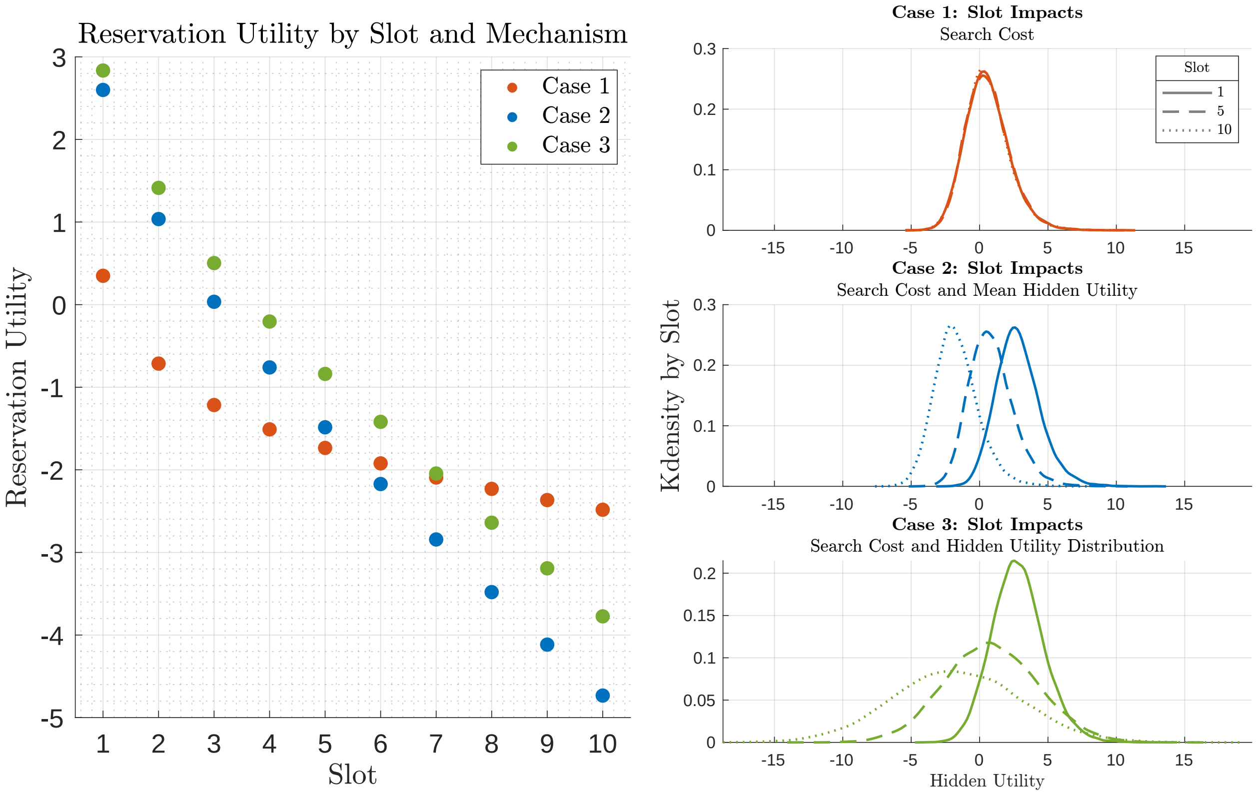
Mechanism 1 is standard in the empirical literature. However, mechanisms 2 and 3 require some additional explanation. Regarding mechanism 2, a consumer might expect that products higher versus lower on the page have different hidden features. In section 1.5, we see the correlation between slot and hidden product features. Another way to think about this is that consumers "trust the algorithm", perceiving products higher on the page as more promising. Regarding mechanism 3, higher-order beliefs, a simplification is that consumers believe the variance of the hidden features to be correlated with slot.[27]

It is important to consider these mechanisms to accurately measure consumer welfare and estimate the correct substitution patterns. Incorrectly excluding one of the mechanisms, may cause in biased results. Consider an example where the true data generating process includes mechanisms 1 and 2, so slot would impact demand both through search cost and rational expectations, but the model only included the search cost mechanism. Observed clicks and purchases purchases would be concentrated near the top of the page because of search cost

[26] $E_i[price_{ijt}|\Omega_{it}] = (1+\tau_m)price^{headline}_{ijt}$ where τ_m is the market-level median % hidden price difference.

[27] For this iteration of the paper, I do not include this mechanism in the empirical model; however, in work in progress, it will be added.

Figure 1.6.1: Simulation Results of Position Effect Mechanisms



and rational expectations, but the model would only be able to explain the pattern in the data through search cost. As a results we would overestimate search cost.

These mechanisms also highlight a limitation of A/B tests and randomized data. If consumers have accurate beliefs about the relationship between slot and hidden utility, there still exists an endogeneity concern with the randomly ordered data. The randomization does address collinearity between features and search costs, but consumers do not know they are in a random treatment group. As a result, their search and purchase decisions would still depend on their beliefs about the recommendation system, i.e., mechanisms 2 and 3.

Accounting for each of these mechanisms requires making different structural assumptions. To validate the assumptions used in the estimation, I re-estimate the demand under alternative assumptions and compare the in- and out-of-sample likelihoods. The results of this exercise are presented in section 1.7.1.7.

1.6.1.11 Characteristics Impacting Reservation Utility

The model also accommodates characteristics that impact demand through reservation utility but do not affect search cost or utility. For example, a promotion on a property, visible on the landing page, certainly impacts utility through the lower price. However, once we control for the price difference, the promotion itself would not change the desirability of the property unless we have a model where consumers derive pleasure from finding "a deal." Nevertheless, promotions may be correlated with hidden product features and thus enter the reservation utility as part of Ω_{it} in $E_i[\delta^h_{ijt}|\Omega_{it}]$.

Similarly, as advertisements or opaque offers may occasionally displace products on the page, I have two variables to keep track of the position on the page: $slot^{appear}_{ijt}$, which denotes where the product appears on the page, and $slot^{rank}_{ijt}$, which denotes the product ranking among non-advertisement products. While $slot^{appear}_{ijt}$ enters the search cost, $slot^{rank}_{ijt}$ enters Ω_{it}, which impacts reservation utility but not utility or the search cost. For example, a product ranked fifth might be displaced by an advertisement in the fifth slot and appear in the sixth slot.

1.6.2 Platform Model of Product Recommendation Systems

This section outlines the platform model, which relates to Expedia's default recommendation system. The platform model and estimates are a necessary component of the structural model to estimate marginal costs on the supply side (hotels) and to provide a baseline for the counterfactual simulations. As discussed in section 1.5, product rankings (slots) play a pivotal role in consumer decision-making. Expedia's recommendation system assigns slots based on query, consumer, and product features, including price. Consequently, firms aiming to maximize profits encounter an elasticity of demand influenced not only by consumer preferences but also by the platform's design. Specifically, a change in a product's price can shift its position in the search results when consumers look for hotels.

The platform model aims to reflect Expedia's default recommendation system accurately. The goal is to generate product recommendations based on a set of queries that match the ranking probabilities of the actual recommendation system and also to capture how a price adjustment for a product alters its likely position in the search results. To achieve this, I use a "model-of-a-model" approach from machine learning to reverse-engineer Expedia's recommendations system.

Expanding on the recommendation system's mechanics, we can write the recommendation systems set up in the format of a demand model with indirect utility but instead thinking of u^r_{ijts} as product j's relevance score for slot s in consumer i's searching at time t:

$$u^r_{ijts} = \omega_s \psi_{ijt} + \epsilon_{ijt} \tag{1.6.13}$$

where u^r_{ijts} denotes the slot-s relevance score of product j for consumer i's query at time

t. $\psi_{ijt} = f(x^r_{ijt})$ defines the deterministic portion of the relevance score, which depends on x^r_{ijt}, a set of consumer, product, and query features. The deterministic score is scaled by ω_s for each slot s. The relevance score also includes some experimental noise, ϵ_{ijt}, which follows a type-1 extreme value distribution. The scale term ω_s is slot-specific since the underlying recommendation system may be relatively more deterministic for some slots than for others. Another way to think of this is that ψ_{ijt} is a seller's (hotel's) expected relevance score conditional on x^r_{ijt}, the information available to the hotel, and ε_{ijt} is the error term.[28]

While this formulation bears similarities to the setup of traditional demand models, there are notable differences. First, an algorithm determines the product recommendations, so the objective of the platform model is to back out the preferences of a single, sophisticated machine. Second, in a demand model, one might expect data on revealed preferences to take the form of clicks and purchases. In the platform model data, the revealed preferences of the algorithm are the complete list of first-page rankings.

In modeling the platform's product recommendation system, it is important to consider that the underlying recommendation system can be a complicated black-box. E-commerce platforms devote significant resources to developing their recommendation systems, using historical data and learning-to-rank methods including neural networks (Ranknet), collaborative filtering (matrix factorization), and gradient-boosted machines (LambdaMART). These methods introduce nonlinearities and high-dimensional interactions, making a straightforward parametric approach prone to misspecification.

Instead of a parametric approach, such as rank-ordered logit, I use a "model-of-a-model" approach from machine learning, also known as model extraction. A growing body of work in the computer science, machine-learning, and cryptography literature demonstrates cases where black-box machine-learning algorithms can be reverse-engineered by training a new model on data generated from queries to the black-box model and the black-box model's results (Papernot et al. (2017); Orekondy et al. (2020)).

In section 1.7.2.2, I outline the estimation procedure to generate the platform model. The estimation procedure involves two steps. In step one, I estimate the function $\psi_{ijt} = f(x^r_{ijt})$ using LambdaMART, a machine learning algorithm used for ranking Burges (2010). Next, I make out-of-fold predictions, $\hat{\psi}_{ijt}$ and estimate the scale terms ω_s for each slot using conditional logit.

[28]Sometimes platforms add noise to the rankings. This can provide useful variation for training future recommendation systems and can prevent price-undercutting strategies in which firms move their price to one cent lower than a competing firm's to move up the ranking.

1.6.3 Supply-Side Model of Hotel Pricing

This section describes the supply side, where capacity-constrained hotels set price schedules and consider how changes in price impact position on the page in search results. I identify hotels in the data based on a property identifier. I treat each hotel as operating independently since I do not observe the hotels' ownership structure.

I focus on the hotels' pricing decisions. However, it is worth noting that there are several decisions hotels can make. They decide prices, can activate promotions on Expedia, and decide whether to sell through Expedia or other platforms such as Booking. They can also choose what percent of the final transaction price to hide from the search results. On the supply side, I hold these decisions fixed, with firms selecting only prices.[29] With these limitations in mind, I model the seller as setting prices to maximize expected profits:

$$\underset{p_{jtt'}}{\operatorname{argmax}} E\left[((1-\varphi)p_{jtt'} - c_{jtt'})q_{jtt'} \mid \Omega_{jtt'}\right] \tag{1.6.14}$$

where $p_{jtt'}$ is the price for room–night j, staying period t, and searching in period t'. φ is the percentage of revenue that goes to the platform and taxes, assuming that both taxes and platform fees are a percent of gross booking revenue. $c_{jtt'}$ denotes the average variable cost. $q_{jtt'}$ is the expected quantity purchased through Expedia; and $\Omega_{jtt'}$ is a hotel's information set, including the own costs, demand elasticities, the features and availability of other products in the same market, and market size. For market size, I assume that the arrival rate of consumers to Expedia is known to firms.

1.6.3.1 Opportunity Cost Interpretation of Average Variable Cost

The firm's problem depends on the average variable cost, $c_{jtt'}$. For this model, $c_{jtt'}$ is a reduced-form object; it is helpful to discuss its interpretation. The hotel sets prices and faces capacity constraints but risks the room–night remaining vacant if it does not sell it by the time of the stay ($t' \geq t$). The interpretation of $c_{jtt'}$ is as the opportunity cost of having room–night jt available to sell in period $t' + 1$.[30]

[29] I conduct the modeling in terms of final price and keep the ratio of hidden-to-visible prices fixed.

[30] Betancourt et al. (2022) details a similar dynamic game for the airline industry. A few limitations prevent me from fully modeling the dynamic game. For example, Betancourt et al. (2022) focuses on two competing airlines and observes quantity and capacity. In contrast, I observe a random sample of quantity but not the capacity constraint, and I have hundreds of competing hotels in the market.

1.6.3.2 Ancillary Revenue

Another aspect of the hotel industry worth noting is that hotels make additional profits post-booking. Hotels' additional goods and services such as room service, bars, and restaurants generate ancillary revenue. In the extreme, we might think about hotel-casinos, where the rooms themselves can be a loss leader, with profits coming from the casino. These ancillary revenue streams are embedded within the $c_{jtt'}$ value.

1.6.3.3 Economies of Scale and Capacity Constraints

In the accommodation industry, average variable costs depend on quantity. Hotels may face economies of scale at low quantities, as adding another guest may not require additional staffing. At high quantities, hotels may face increasing marginal costs, for example, from needing to pay overtime to meet staffing requirements. Further, hotels face capacity constraints, which imply increasing opportunity cost as quantity approaches the capacity constraint. I capture these features by allowing average and marginal costs to depend on quantity and quantity squared. Equations 1.6.15 and 1.6.16 specify the marginal and average variable costs functions in relation to quantity. These costs exclude any of the large fixed costs typical of the hotel industry.

Cost Functions

$$\text{average cost:} \quad c_{jtt'} = mc^{\text{base}}_{jtt'} + \frac{1}{2}\gamma_{1jtt'}q_{jtt'} + \frac{1}{3}\gamma_{2jtt'}q^2_{jtt'} \tag{1.6.15}$$

$$\text{marginal cost:} \quad mc_{jtt'} = mc^{\text{base}}_{jtt'} + \gamma_{1jtt'}q_{jtt'} + \gamma_{2jtt'}q^2_{jtt'} \tag{1.6.16}$$

where $mc^{\text{base}}_{jtt'}$ is the variable cost associated with providing one additional unit of accommodation before considering any effects from economies of scale or increasing costs due to higher occupancy levels. This represents the base per-unit cost of accommodating a guest for hotel j, for a stay at time t, and at search time t'. $\gamma_{1jtt'}$ is a negative relationship between marginal cost and quantity that captures economies of scale. $\gamma_{2jtt'}$ is a positive term on quantity squared that captures increasing costs and serves as a soft capacity constraint.

This approach is similar to that in Farronato and Fradkin (2022), which models hotel capacity constraints with a hockey stick–type function, with flat cost for low quantity and then linearly increasing cost above 85% occupancy. I do not observe capacity; however, the polynomial specification in terms of quantity should be able to capture the inflection point where marginal costs increase. In the primary specification, γ_1 and γ_2 are star-rating specific. However, with enough data, one could specify a firm-specific γ_1 and γ_2.

1.6.3.4 Seller First-Order Condition

We can take the derivative of the seller's problem with respect to $p_{jtt'}$ to obtain the profit-maximizing first-order condition:

$$\frac{mc_{jtt'}}{(1-\varphi)} = p_{jtt'} + \left(\frac{\partial q_{jtt'}}{\partial p_{jtt'}}\right)^{-1} q_{jtt'} \tag{1.6.17}$$

I do not observe the percent of revenue that goes to taxes and fees, φ, so I express the marginal cost as a ratio of $(1-\varphi)$. This is not an issue for estimation as long as φ remains fixed.

If we allow marginal cost to depend on quantity, the numerator of the left-hand side of the problem becomes $mc_{jtt'} = mc^{\text{base}}_{jtt'} + \frac{\partial c_{jtt'}}{\partial q_{jtt'}} q_{jtt'}$. In section 1.7.3, I discuss estimating costs. Since costs depend on quantity, I need an additional model of cost and instruments for quantity and quantity squared.

1.7 Estimation and Results

This section discusses the estimation procedure and results for the empirical model presented in the previous section. I estimate the demand and platform models separately. I then use their combined results to estimate the supply-side model.

1.7.1 Demand Estimation and Results

I use the optimal sequential search rules from Weitzman (1979) and logit smoothing techniques covered in Train (2009) and proposed in McFadden (1989) to construct the joint likelihood of clicking and booking decisions.

1.7.1.1 Utility Specification

Writing out the primary per-night utility specification, we have:[31]

$$\text{inside option: } u_{ijtt'} = \underbrace{\beta_i^v x^v_{jtt'} + \xi_t^{\text{month}} + \xi_t^{\text{day}}}_{\delta^v_{ijt}} \underbrace{-e^{\rho_i} \overbrace{p_{jtt'}}^{\text{final price}} + \beta_i^h x_j^h}_{\delta^h_{ijt}} + \underbrace{\lambda \varepsilon^v_{ijtt'} + \varepsilon^h_{ijtt'}(\lambda)}_{\text{match quality}\sim EV1} \tag{1.7.1}$$

$$\text{outside option: } u_{i0tt'} = \alpha_0 + \varepsilon_{i0tt'} \tag{1.7.2}$$

[31] For the rest of this section, I again drop the tt' subscript.

where x_j^v are the visible product features, p_{jt} is the final transaction price (including taxes and fees), and x_j^h are the hidden product features. The price coefficient, e^{ρ_i}, follows a log normal distribution. x_j^v includes indicators for star-rating, an interaction on brand and star rating, a linear spline of review score, indicators for missing values, and consumer segment information. The consumer segments are quantiles based on the booking window of the stay, the time of the search (morning, working hours, evening, and weekend or weekday), and the length of stay. The hidden features include splines on both location desirability scores and a missing indicator for location desirability score 2. The time effects ξ_t^{month} and ξ_t^{day} control for the market–month and market–day of the week of the stay, respectively. The final price p_{jt} is assumed to be hidden, as consumers see the headline price on the landing page.

The utility specification includes random coefficients on the indicators for star rating, brand, and price. The price coefficient is correlated with search costs. I estimate the elements of the Cholesky decomposition of the random coefficient covariance matrix.

To save on notation, I use the following expression for within-estimation utility:

$$u_{ijt}^{[s]} = \delta_{ijt}^{v[s]} + \delta_{ijt}^{h[s]} + \lambda \varepsilon_{ijt}^{v[s]} + \varepsilon_{ijt}^{h[s]}(\lambda) \tag{1.7.3}$$

This notation adds a superscript $[s]$ that indexes the set of draws, so $u_{ijt}^{[s]}$ means that this is simulated per-night utility for consumer i, product j, for stay at time t, (searching at time t'), and simulated draws s. The draws are scrambled-Halton draws for utility parameters, the random coefficients, and the error terms.

The hidden match quality term, $\varepsilon_{ijt}^{h[s]}(\lambda)$, depends on the parameters λ. Each iteration of the estimation loop requires producing a new $\varepsilon_{ijt}^{h[s]}(\lambda)$. Prior to estimation, I take Halton draws that do not change during estimation, then use the approximate Cardell distribution to obtain $\varepsilon_{ijt}^{h[s]}(\lambda) = \text{ICDF}(\lambda, d_{ijt}^{[s]})$.

1.7.1.2 Search Cost

The search cost follows from the model parameters and $\text{slot}_{ijt}^{\text{appear}}$.

$$c_{ijt}^{[s]} = \log\left(1 + \exp\left(\kappa_i^{[s]} + \sum_{k \in K} \tau_k \left(\log\left(\text{slot}_{ijt}^{\text{appear}}\right) - \gamma_k\right)_+\right)\right) \tag{1.7.4}$$

The log-exponential functional form above guarantees positive search costs. $\kappa_i^{[s]}$ is normally distributed with mean κ and correlated with the price coefficient ρ_i. The position on the page $\text{slot}_{ijt}^{\text{appear}}$ enters search cost with a spline function. Using splines on slot allows me to flexibly capture the relationship between search cost and slot. This has the added benefit of making the results relatively robust to the functional form assumption for search cost.

1.7.1.3 Heterogeneous Preferences and Search Costs

The demand model allows for a rich set of random coefficients on utility and search cost. The primary specification includes random coefficients on price, the inside option, star rating (1–5), and search cost. Additionally, the primary specification includes correlated random coefficients on search cost and price.

The star ratings indicate the hotel class.[32] The random coefficient on each star rating serves a similar purpose to nests in a nested logit, where consumers have correlated tastes for hotels within the same class.[33] An extension of the model would allow for correlation among the star-rating coefficients.

The random coefficient on price and search cost allows different consumers to have different search costs and different price sensitivities. Allowing for price–search cost correlation is sensible, as one interpretation of the search cost is that it is partially the opportunity cost of time, and the price parameter captures the opportunity cost of money. For example, a high income consumer might have a high opportunity cost of time and a low opportunity of money.

1.7.1.4 Reservation Utility

Next, the reservation utility consists of four elements:

$$r_{ijt}^{[s]} = \delta_{ijt}^{v[s]} + E[\delta_{ijt}^{h[s]}|\Omega_{it}] + \zeta_{ijt}^{[s]} + \lambda\varepsilon_{ijt}^{v[s]} \tag{1.7.5}$$

Utility from visible features, $\delta_{ijt}^{v[s]}$, the visible match quality, $\lambda\varepsilon_{ijt}^{v[s]}$, the expected utility from hidden features $E[\delta_{ijt}^{h[s]}|\Omega_{it}]$, and $\zeta_{ijt}^{[s]}$, which corresponds to the portion of reservation utility that satisfies equation 1.6.9. $\delta_{ijt}^{v[s]}$ and $\lambda\varepsilon_{ijt}^{v[s]}$ can be calculated directly from the model parameters, consumer–product–draw features and random draws. $E[\delta_{ijt}^{h[s]}|\Omega_{it}]$ and $\zeta_{ijt}^{[s]}$ require additional processing.

Expected Hidden Utility One approach to calculating $E[\delta_{ijt}^{h[s]}|\Omega_{it}]$ is to estimate a linear regression with $\delta_{ijt}^{h[s]}$ as the left-hand side and the relevant variables from Ω_{it} on the right-hand side and then predict the values of $\delta_{ijt}^{h[s]}$. This would be computationally burdensome, as $\delta_{ijt}^{h[s]}$ depends on the model parameters and so this estimation and prediction would need to be repeated with each evolution of the objective function. Alternatively, to save estimation time, I use linearity of expectations to express expected hidden utility $E[\delta_{ijt}^{h[s]}|\Omega_{it}]$ as a function of

[32] https://www.expedia.com/Hotel-Star-Rating-Information

[33] Train (2009) discusses the similarity between a model with a random coefficient on a categorical variable and a nested logit with a category nest.

model parameters and expected hidden features:

$$E[\delta_{ijt}^{h[s]}|\Omega_{it}] = -e^{\rho_i^{[s]}} E[p_{ijt}|\Omega_{it}] + \beta_{ijt}^{h[s]} E[x_{ijt}^{h[s]}|\Omega_{it}] \tag{1.7.6}$$

Since the expected features can be estimated directly from data (with rational expectations), I estimate $E[p_{ijt}|\Omega_{it}]$ and $E[x_{ijt}^{h[s]}|\Omega_{it}]$ outside the estimation loop. The expected final price, $E[p_{ijt}|\Omega_{it}]$, is the headline price of the hotel multiplied by the median hidden price percentage for the market. For the features, $E[x_{ijt}^{h[s]}|\Omega_{it}]$, consumers know the star rating, whether the hotel is on promotion, and a spline of the logged $\text{slot}_{ijt}^{\text{rank}}$.

Value Function Approximation The final component of reservation utility, $\zeta_{ijt}^{[s]}$, does not have an analytic expression but can be solved numerically, as I know that reservation utilities satisfy the equality in equation 1.6.9. Numerically solving for each $\zeta_{ijt}^{[s]}$ for every evaluation of the objective function would obviously be computationally infeasible. Instead, researchers solve for $\zeta_{ijt}^{[s]}$ numerically, on a fine grid of state variables, and then use curve fitting to estimate $\zeta_{ijt}^{[s]}$ not exactly at the grid points. This is a point on which my approach differs from common approaches in the search literature. Kim et al. (2010) establish a commonly used approach where $\zeta_{ijt}^{[s]}$ can be solved numerically prior to estimation on an arbitrarily fine grid (this approach relies on the assumption that 1) consumers know product features, and only learn match quality from search).[34] However, since consumers learn about product features and the λ term, $\zeta_{ijt}^{[s]}$ depends on too many parameters for $\zeta_{ijt}^{[s]}$ to be feasibly solved outside the estimation loop. Instead, I move the value function approximation that yields $\zeta_{ijt}^{[s]}$ inside the estimation loop and use a grid interpolation approach,[35] following a common approach used in economics to approximate value functions.

To do this, I include an inner loop, where I numerically solve for the ζ component on a grid of state variables then fit a spline interpolation object to the grid. ζ depends on all the hidden utility parameters, search cost, price coefficients, λ, length of stay, and the random coefficients on hidden features and price. In the primary specification, the grid includes 1,692 point.

Position Variables The position variables, $\text{slot}_{ijt}^{\text{rank}}$ and $\text{slot}_{ijt}^{\text{appear}}$, impact reservation utility differently. $\text{slot}_{ijt}^{\text{rank}}$ is used to shift the expectation of hidden features through $E[\delta_{ijt}^{h[s]}|\Omega_{it}]$. In the current setup, $\zeta_{ijt}^{[s]}$ varies with $\text{slot}_{ijt}^{\text{appear}}$ through search costs. As an extension, it would be possible to account for higher-order consumer beliefs about the relationship between product

[34]Other examples that use this approach include Chen and Yao (2017) and Ursu (2018).
[35]Examples of grid interpolation appear here:
https://uk.mathworks.com/help/matlab/ref/griddedinterpolant.html.

rankings and hidden features by including $\text{slot}_{ijt}^{\text{rank}}$ as a state variable in the grid interpolation object. Of course, this comes with a practical tradeoff of requiring more points to solve for ζ numerically.

Optimal Sequential Search Setup

The optimal sequential search model from Weitzman (1979) establishes four rules: order, continuation, stopping, and choice. These four conditions can collectively identify the ordered search and purchase decisions of consumers.

In this section, I go through each rule in turn, consolidate the rules as applying to clicks or purchases, and then use logit smoothing to convert these rules into joint likelihoods.

In this setup, it is helpful to define an index of consumer actions, m, where m_{it}^* refers to the last action (which is always a purchase or choice of the outside option) and each $m < m_{it}^*$ refers to a click. We can use this index to identify observed consideration sets $S_{it}(m)$, which refer to the first $m-1$ items clicked on by consumer i, and the outside option. Going through each rule in turn, we have:

Order Rule: Consumers search in descending order of reservation utility. This means that, at any stage in the search process, the next-clicked item must have a higher reservation utility than that of all of the not-clicked and not-yet-clicked items. Writing out the condition, we have

$$\forall m < m_{it}^*, \quad r_{ijt} \geq r_{ikt} \; \forall k \notin S_{it}(m) \tag{1.7.7}$$

where m is the m-th step of the search process and m_{it}^* is the number of clicks for consumer search it. $S_{it}(m)$ is consumer i's ordered consideration set of $m-1$ already searched items and the outside options.

Continuation Rule Search continues if any unsearched items have a higher reservation utility than the best option in the consideration set. Formally,

$$\forall m < m_{it}^*, \exists k^* \notin S_{it}(m) \colon r_{ik^*t} \geq u_{ikt} \quad \forall k \in S_{it}(m) \tag{1.7.8}$$

Stopping Rule Search stops if the utility from the best option so far (including the outside option) is greater than the reservation utilities of all the remaining unsearched options.

$$\exists k^* \in S_{it}(m_{it}^*) \colon u_{ik^*t} \geq r_{ikt} \quad \forall k \notin S_{it}(m_{it}^*) \tag{1.7.9}$$

where $S_{it}(m_{it}^*) = S_{it}$ is the consumer's complete ordered-consideration set.

Choice Rule Once search ends, the consumer chooses the product with the highest utility in the consideration set.

$$u_{ijt} \geq u_{ikt} \quad \forall k \in S_{it}(m_{it}^*) \tag{1.7.10}$$

While optimal sequential search has four conditions, in my setting, there are only two types of actions: clicks and purchases (including choosing the outside option). The order and continuation rules apply to clicks, and the stopping and choice rules apply to purchases. Combining the rules, we have the following two consolidated conditions.

Click Rule

The click rule combines the order and continuation conditions.

Formally, we can combine them as:

$$\forall m < m_{it}^*, \quad \underbrace{(r_{ijt} \geq r_{ik't} \quad \forall k' \notin S_{it}(m))}_{\text{order rule}} \wedge \underbrace{(\exists k^* \notin S_{it}(m)\colon r_{ik^*t} \geq u_{ikt} \quad \forall k \in S_{it}(m))}_{\text{continuation rule}} \tag{1.7.11}$$

For the m-th clicked item, it must have the highest reservation utility among the not-(yet-)searched items and also have a higher reservation utility than the items already in the consideration set. If the item has a lower reservation utility than a different product, then that product would be clicked instead. If the product has a lower reservation utility than a product already in the consideration set, then the search would stop.

We can further simplify the click rule as follows:

$$\forall m \leq m_{it}^*, (r_{ijt} \geq u_{ikt} \quad \forall k \in S_{it}(m)) \wedge (r_{ijt} \geq r_{ik't} \quad \forall k' \notin S_{it}(m)) \tag{1.7.12}$$

Purchase Rule

The purchase rule combines the stopping and choice conditions. For each consumer, the purchase rule applies only to the last action.

$$\underbrace{(\exists k^* \in S_{it}(m_{it}^*)\colon u_{ik^*t} \geq r_{ikt} \quad \forall k \notin S_{it}(m_{it}^*))}_{\text{stopping rule}} \wedge \underbrace{(u_{ijt} \geq u_{ikt} \quad \forall k \in S_{it}(m_{it}^*))}_{\text{choice rule}} \tag{1.7.13}$$

This is equivalent to saying that the chosen item (in step m_{it}^*) action satisfies the purchase rule if the chosen product has a higher utility than that of all other items in the consideration set and a higher utility than the reservation utilities of the not-searched items.

Joint Likelihood Construction and Logit Smoothing

Now that we have the consolidated click rules and purchase rules, we can construct the joint likelihood of sequential search and purchase decisions. One approach would be to use an accept–reject (AR) simulator. In the AR simulator, for each consumer-draw ($i[s]$), record a one if each click rule and the purchase rule are satisfied; then, take the average over simulations to obtain the joint likelihood. However, as noted in Ursu (2018)), the dimensionality of ordered sets makes this type of AR simulation impractical; for example, with just ten products, there are over 60 million possible ordered consideration sets and choices. An alternative approach is to use logit smoothing, following Train (2009).[36]

As stated above, I am integrating over the match quality terms, ε^h and ε^v, with scrambled Halton sequences. In logit smoothing, I conduct the modeling as if there is a type-1 extreme value term, scaled by a smoothing parameter ω, associated with each click and choice condition. I can then obtain a logit-smoothed expression for the click and purchase conditions.

Logit-Smoothed Click-Condition

$$P_{it}^{\text{click[s]}} = \prod_{m \in S_{it}} \left(\frac{\exp\left(\frac{x_{it}^{\text{nights}} r_{imt}^{[s]}}{\omega}\right)}{\sum_{k \in S_{it}(m)} \exp\left(\frac{x_{it}^{\text{nights}} u_{ikt}^{[s]}}{\omega}\right) + \sum_{k' \notin S_{it}(m)} \exp\left(\frac{x_{it}^{\text{nights}} r_{ik't}^{[s]}}{\omega}\right)} \right) \tag{1.7.14}$$

where $P_{it}^{\text{click[s]}}$ is the smoothed likelihood that all $m_{it}^* - 1$ clicks for consumer i satisfy the click conditions. ω denotes the smoothing parameter, and x_{it}^{nights} is the length of stay.

Logit-Smoothed Purchase Condition

$$P_{it}^{\text{choice[s]}} = \frac{\exp\left(\frac{x_{it}^{\text{nights}} u_{ijt}^{[s]}}{\omega}\right)}{\sum_{k \in S_{it}} \exp\left(\frac{x_{it}^{\text{nights}} u_{ikt}^{[s]}}{\omega}\right) + \sum_{k' \notin S_{it}} \exp\left(\frac{x_{it}^{\text{nights}} r_{ik't}^{[s]}}{\omega}\right)} \tag{1.7.15}$$

where $P_{it}^{\text{click[s]}}$ is the smoothed likelihood that consumer i's purchase decision satisfies the purchase conditions.

[36] Logit smoothing is a popular approach to non–smooth objective functions. It is also a useful tool for search models for example Honka (2014), Ursu (2018), and Honka and Chintagunta (2017) use various logit-smoothing Logit smoothing is also a well-established technique in computer science for smoothing loss functions, in which ω is referred to as the temperature parameter (Platt, 2000)

Joint Likelihood

We can now combine these conditions to obtain the logit-smoothed joint likelihood of search and purchase decisions.

$$\mathrm{P}_{it}^{[s]} = \overbrace{\prod_{m \in S_{it}} \underbrace{\left(\frac{\exp\left(\frac{x_{it}^{\text{nights}} r_{imt}^{[s]}}{\omega}\right)}{\sum_{k \in S_{it}(m)} \exp\left(\frac{x_{it}^{\text{nights}} u_{ikt}^{[s]}}{\omega}\right) + \sum_{k' \notin S_{it}(m)} \exp\left(\frac{x_{it}^{\text{nights}} r_{ik't}^{[s]}}{\omega}\right)} \right)}_{\text{click condition for } m\text{-th click}}}^{\text{click condition(s)}} \times \overbrace{\left(\frac{\exp\left(\frac{x_{it}^{\text{nights}} u_{ijt}^{[s]}}{\omega}\right)}{\sum_{k \in S_{it}} \exp\left(\frac{x_{it}^{\text{nights}} u_{ikt}^{[s]}}{\omega}\right) + \sum_{k' \notin S_{it}} \exp\left(\frac{x_{it}^{\text{nights}} r_{ik't}^{[s]}}{\omega}\right)} \right)}^{\text{purchase condition}} \tag{1.7.16}$$

where $P_{it}^{[s]}$ is the joint likelihood of the observed search and purchase decisions for consumer i in search t within the set of simulated draws s. Averaging $P_{it}^{[s]}$ over simulations yields the likelihood:

$$P_{it}^{\text{joint}} = \frac{1}{D} \sum_{s=1}^{D} P_{it}^{[s]} \tag{1.7.17}$$

where D denotes the number of simulations (draws).

1.7.1.5 Sample Selection Adjustments

As discussed in section A.1.3, there are two types of sample selection. These data include only observations with at least one click, and observations with a purchase are oversampled relative to searches without a purchase. Not adjusting for these sampling issues would lead to biased parameter estimates. I adjust for the first (any click) by using conditional likelihoods. I adjust for the second issue by using observation weights.

Conditioning on Any Click

The sample includes only data from searches in which consumers clicked at least one of the options. I adjust for this selection by conditioning the likelihoods on clicking at least one item. This requires calculating the likelihood of clicking at least one item. At the consumer it–simulation $[s]$, level we can express this as the likelihood of at least one reservation utility being greater than the utility of the outside option.

Smoothed likelihood of any clicks. The smoothed likelihood of making any clicks for consumer i at time t for simulation s is given by

$$P_{it}^{\text{any click[s]}} = \frac{\sum_k \exp\left(\frac{x_{it}^{\text{nights}} r_{ik't}^{[s]}}{\omega}\right)}{\exp\left(\frac{x_{it}^{\text{nights}} u_{i0t}^{[s]}}{\omega}\right) + \sum_k \exp\left(\frac{x_{it}^{\text{nights}} r_{ik't}^{[s]}}{\omega}\right)} \tag{1.7.18}$$

It is possible to estimate $P_{it}^{\text{any click[s]}}$ without smoothing; however, as it is used to condition the joint likelihood $P_{it}^{[s]}$, which is smoothed, I use the same smoothing approach for $P_{it}^{\text{any click[s]}}$, which avoids conditional likelihoods above 1.

Conditional likelihood. Using the definition of conditional probability, we can write the conditional likelihood of a consumer's ordered search and purchase conditional on clicking at least one hotel as

$$P_{it|\text{any click}}^{\text{joint}} = \frac{\frac{1}{D}\sum_{s=1}^{D} P_{it}^{[s]}}{\frac{1}{D}\sum_{s=1}^{D} P_{it}^{\text{any click[s]}}} \tag{1.7.19}$$

where $P_{it}^{[s]}$ is the unconditional joint likelihood of search and purchase from 1.7.16 and $P_{it}^{\text{any click[s]}}$ is the likelihood of any clicks from 1.7.18.

Sample Weights

To address the different sample rate for observations with versus without a purchase, I use sample weights to achieve a target purchase rate (conditional on at least one click) of 16.66%. I calculate the weights using data from the top 5 markets. The relative weight depends on the observed purchase decisions

$$w_i = \begin{cases} w^{\text{in}}, & \text{if consumer } i \text{ chose an inside good} \\ w^{\text{out}}, & \text{otherwise} \end{cases} \tag{1.7.20}$$

where w_i is the weight for consumer i (note that I do not observe consumer IDs across searches). In the primary specification, w^{in} is normalized to one and $w^{\text{out}} = 56.63$.

Ignoring the sampling issue would lead to biased parameters. The direction of some parameters is not obvious, but a simple way of thinking about this problem is the following. In the observational data, there is approximately a 90% conversion rate. Without weights, the inside option would seem highly desirable when, in reality, consumers rarely make a purchase. Ignoring the weighting would also cause other concerns, as match quality terms and random coefficients would also influence the decision to make the first click.

Weighted Log Simulated Likelihood

Applying sample weights and using the conditional likelihoods for each observation yields the consumer–search-level weighted likelihood:

$$wll_i = w_i \log\left(\frac{1}{D}\sum_{s=1}^{D} P_{it}^{[s]}\right) - w_i \log\left(\frac{1}{D}\sum_{s=1}^{D} P_{it}^{\text{any click[s]}}\right) \tag{1.7.21}$$

Summing across consumers yields the logit-smoothed log simulated likelihood.

$$SLL = \sum_i wll_i \tag{1.7.22}$$

Algorithm 1 in Appendix A.5.0.1 summarizes the demand estimation procedure.

1.7.1.6 Informal Identification

Since this is a maximum simulated likelihood estimation, to some extent, everything identifies everything. However, it is helpful to discuss the intuition for parameter identification in terms of the optimal sequential search rules and notable variation in the data. Table 1.7.1 summarizes the key sources of identification, and I discuss further details below.

Table 1.7.1: Informal Identification of Demand Parameters

	Sequential Search Conditions				Notable Variation		
Parameters	Order	Continuation	Stopping	Choice	Nights	Diversion	Displacement
Utility Parameters							
Consumer Segments: δ_{it}		✓	✓	✓†	✓	✓	✓
Time Effects: ξ_{it}^{month},ξ_{it}^{day}		✓	✓	✓†	✓	✓	✓
Mean: ρ, β^v, β^h	✓	✓	✓	✓	✓	✓	✓
Heterogeneous: Σ_u	✓*	✓*	✓*	✓*	✓	✓	✓
Visible Error Scale: λ	✓	✓	✓		✓	✓	✓
Search Cost Parameters							
Mean: κ, τ_k	✓	✓	✓		✓	✓	✓
Heterogeneous: Σ_κ	✓*	✓*	✓*		✓	✓	✓

Note: Checkmarks with an asterisk (✓*) indicate parameters that are identified by repeated decisions within consumer (e.g., clicks and purchase). Checkmarks with a dagger (✓†) indicate parameters that are identified by selecting an inside good versus the outside option, but not from the choice of one inside good over another. "Nights" refers to length of stay. "Diversion" refers to substitution patters from variation in product features and availability. "Displacement" refers the variation in positions caused by advertisements/opaque offers.

Length of stay separately identifies search costs and utility parameters. The challenge of separately identifying search cost and utility parameters in search models is well-documented. For example, slot affects search costs but is often highly correlated with product features. Koulayev (2014) notes the challenge of distinguishing high search costs from low tastes.[37] My approach addresses this issue by leveraging variation in length of stay. Consumers searching for longer stays would be consuming more of the good and paying a multiple of the prices. This means returns to search depend on length of stay. However, length of stay is, presumably, independent of search costs. More formally, two consumers with identical utility and search cost parameters but different lengths of stay would have the same per-night utilities, but different reservation utilities. To my knowledge, this is the first paper to take advantage of length stay to address these identification issues.

Diversion and displacement. Other variation also helps me separately identify utility and search costs parameters, including diversion (similar to the diversion ratio), where I observe different search and purchase decisions under different hotel availability, product rankings and prices. Additionally, in some searches, opaque offers appear and displace the positioning of some hotels; this means that in some scenarios, $slot_{ijt}^{rank} \neq slot_{ijt}^{appear}$.

Repeated within-consumer decisions. Although I do not observe multiple search sessions for each consumer, I still observe repeated decisions within a search session. A consumer's clicks and purchase within a session help to identify random coefficients. For example, if consumers click only on hotels of the same star rating, that is indicative of the random coefficients on star rating. Similarly, the correlation of prices within a consumer's consideration set informs the random coefficient on price.

Market–time effects. While it might be reasonable to assume that the hotels can be modeled in feature space, a key feature of the accommodation industry is that prices move with time-varying demand. For example, in a college town, prices and demand increase during sporting events and graduations. I include market–month and market–day-of-week effects. This assumes that the time-varying shifts in demand occur at the market level and are not hotel–time specific. As an extension, I could use narrower time effects, for example, market–week instead of market–month.

[37]See Ursu et al. (2023) for an overview other challenges, such as cases where parameters that suggest both returns to search are and search costs are high, produce results similar to those where parameters suggest that both returns to search and search costs are low.

Feature space. I model utility for hotels in feature space, relying on the assumption that the rich set of product features captures the hotel-specific utilities. There are over 700 products in the top market, and many of them appear rarely, making a product fixed effect approach impractical.

1.7.1.7 Demand Results

The primary specification of the model uses a 90% sample of observations from the top market, with observations that were subject to Expedia's default recommendation system.[38] Table 1.7.2 presents the parameter estimates.[39] The results are consistent with intuition. The λ parameter takes on a value of 0.28, suggesting that consumers know part of the match quality prior to search but learn most from the search. Four- and five-star hotels have higher mean utility than lower-rated hotels. Search cost is monotonically increasing in page position (this is not a constraint). In include additional results on search costs and implied differences in reservation utility by slot in A.6.

[38]The remaining 10% are used to evaluate the out-of-sample performance of the model.

[39]standard errors are still a work in progress, as bootstrapping this model is computationally expensive

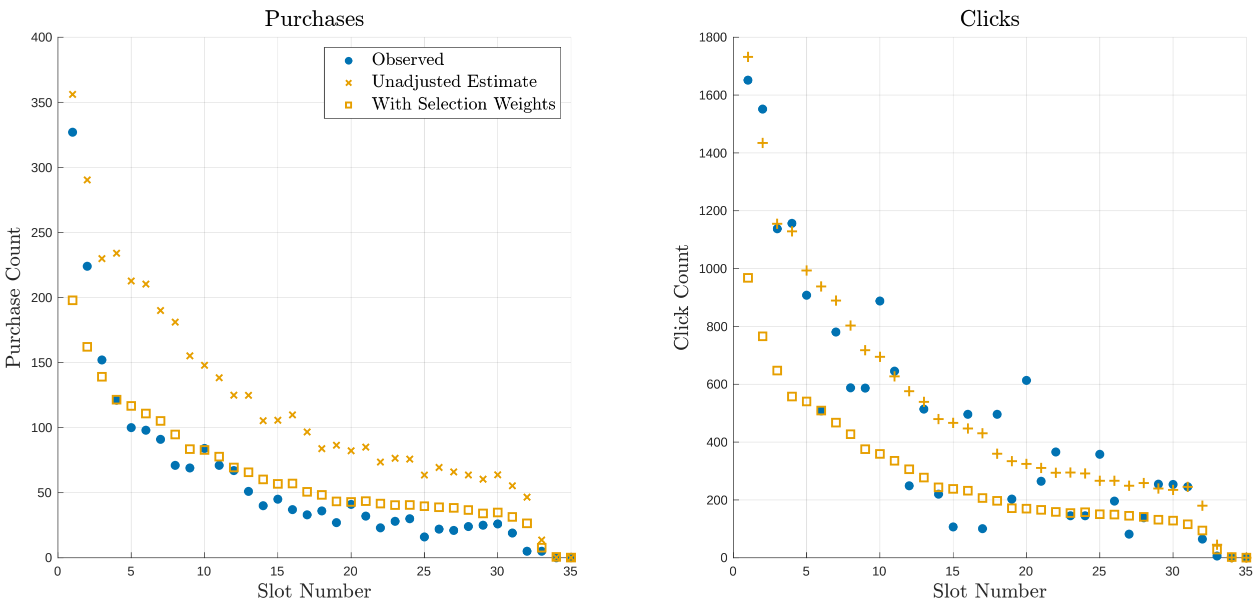
Table 1.7.2: Demand Parameter Estimates

Variable	Coefficient
Match quality split λ	0.28
Outside option	1.90
Price ($100s)	-1.76
3 star	0.30
4 star	0.54
5 star	0.48
Non-star	0.31
2 star brand	-0.16
3 star brand	-0.28
4 star brand	0.03
5 star brand	0.29
Prop review score 1 to 3	-0.51
Prop review score 3 to 5	0.04
Mi. prop review score	-1.40
Hidden: location score 1 (spline 1)	0.52
Hidden: location score 1 (spline 2)	-0.51
Hidden: location score 1 (spline 3)	0.05
Hidden: location score 1 (spline 4)	2.61
Hidden: location score 2 (spline 1)	0.27
Hidden: location score 2 (spline 2)	1.50
Hidden: location score 2 (spline 3)	0.37
Hidden: mi. location score 2	1.64
Search cost: constant	-1.10
Search cost: ln slot spline 1	0.11
Search cost: ln slot spline 2	0.21
Search cost: ln slot spline 3	0.37
Search cost: ln slot spline 4	0.08
Day of week	✓
Month	✓
Time before stay	✓
Length of stay	✓
Search time	✓
Search on weekends	✓
Random coefficients	✓
Correlated price and search cost	✓
Obs	2,262
Obs (weighted)	13,444
Halton draws	400
Smoothing param ω	.2
Grid points	1,692
Log likelihood	-85,027.995

Model Fit

Figure 1.7.1 plots observed purchases (left) and clicks (right) by position on the page. It compares the observed amounts to the expected quantities implied by the parameter estimates from the demand model.

Figure 1.7.1: Demand Fit: Predicted vs Observed Quantity and Clicks



Note: Base on the in-sample data from the top market.

Both figures show the expected pattern of observed and predicted counts of purchases and clicks decreasing as we move down the page. The selection weights adjustment appears to bring the estimated counts closer to the observed counts, but there is still a noticeable gap, especially at lower slot numbers, where the models underestimate the position effects, resulting in lower predicted counts of purchases and clicks compared to the observed levels. This suggests that, while the model has some predictive accuracy, there is room for improvement.

Position Effect Mechanisms: Search Cost and Rational Expectations

Here, I briefly return to the discussion on the mechanisms driving search cost. To test the structural assumption that position impacts demand through search cost and expectation of hidden features, I reestimate the demand using alternative structural assumptions. Table 1.7.3 shows the results, comparing the primary specification results to those of one where position impacts demand only through search cost. The primary specification, which allows position effects to be driven by both search costs and beliefs, outperforms the search cost–only model both in and out of sample.

Table 1.7.3: Position Effect Mechanism Results

	Position Effect Structural Assumption	
	Search Cost	**Search Cost & Beliefs**
Log Likelihood (In Sample)	-85,567	-85,028
Log Likelihood (Out of Sample)	-13,936	-13,914

Note: Logit-smoothed joint likelihoods of search and purchase conditional on at least one click. Includes sampling weights based on conversion rates. 2262 in-sample (training) observations, 251 out-of-sample (testing) observations.

Both models have the same number of underlying parameters. The only difference is the structural assumptions, so we can directly compare the log-likelihoods. If we were testing a model with alternative parameters, then we would need a measure, such as the Akaike or Bayesian information criterion (AIC or BIC), that includes a penalization for additional parameters.

1.7.2 Platform Model: Product Recommendation Model Estimation and Results

This section describes the estimation procedure for the platform model. I estimate the platform model with the naturally ordered Expedia data, using the natural rankings as the outcome variable and the product, query, and consumer features as the explanatory variables. The estimation procedure consists of two steps. The first step uses a "model-of-a-model" approach from machine learning to estimate the deterministic portion of the relevance scores. The second step is to estimate a set of slot-specific conditional logits, which scale the deterministic portion of the relevance scores from the first step.

In section 1.8, I provide a more detailed overview of learning-to-rank methods. In that section, I develop personalized rankers explicitly for ranking, whereas in this section, although I use a learning to ranking method, I do so to model the behavior of Expedia's default recommendation system.

I use a "model-of-a-model" approach from machine learning. This connects to the the growing body of work in the computer science, machine-learning, and cryptography literature that demonstrates cases where black-box machine-learning algorithms can be reverse-engineered by training a new model on data generated from queries to the black-box model and the black-box model's results (Tramer et al., 2016; Papernot et al., 2017; Oh et al., 2018; Hu and Pang, 2021).

1.7.2.1 Platform Model Step 1: Model Extraction via LambdaMART

The first step of the platform model is the model extraction step which applies LambdaMART, a machine learning algorithm used for ranking problems, detailed in Burges (2010), to the naturally ordered data. The name "LambdaMART" is derived from the fact that it is a combination of "Lambda" (referring to the gradient boosting approach it uses, which computes lambda-like quantities) and "MART" (Multiple Additive Regression Trees). LambdaMART has proven to be one of the more effective ranking algorithms, and is a popular choice of algorithm in data science competitions and industry. Microsoft uses LambdaMART as the underlying algorithm in Bing's search engine.[40] An ensemble of LambdaMART rankers won Track 1 of the Yahoo! Learning to Rank Challenge (Chapelle and Chang (2011)), and an ensemble rankers including LambdaMART also won the Personalize Expedia Hotel Searches – ICDM 2013 competition, the source of my data.

Pseudo-Relevance Score

Learning-to-rank models rely on a relevance score, so I convert the product rankings to scores between zero and five, where a five is the top slot, and 0 is the last product listed.

$$rel_{ijt} = 5 - \frac{1 + max_k(slot^{rank}_{ikt}) - slot^{rank}_{ijt}}{max_k(slot^{rank}_{ikt})} \tag{1.7.23}$$

Where rel_{ijt} is the relevance score for product j in consumer i's query at time t. $slot^{rank}_{ijt}$ is the position on the page, adjusting for advertisements and opaque offers.[41] A product in the first slot receives a relevance score of five, while a product in the last slot, $max_k(slot^{rank}_{ikt})$, would receive a relevance score of zero.

Input Data

The input data for the first stage of the platform model include the training observations for the top five markets. I incorporate a range of input variables and interactions. These include product features including headline price, promotion indicator, hidden price percentages, star rating, review score, brand indicators, location scores, search query affinity, the log of the

[40] https://www.microsoft.com/en-us/research/blog/ranknet-a-ranking-retrospective/

[41] If a product is in the sixth slot, but the fifth slot is occupied by an opaque offer, then $slot^{rank}_{ijt}$ would be five, since it is ranked fifth by the recommendation system. This assumes advertisements and opaque offers are determined separately, which is reasonable since they occupy the same position on the page.

historical price and indicators for missing values. I also include product specific data from eight competing OTAs. At the query level, I account for market id, and submarket id, time of search, booking window, weekend searches, if the stay includes a weekend, and specific site indicators. The query level variables do not vary withing a search, so I include these variables as interactions with the product features variables listed above.

Step 1 Estimation

In this setup, the features are the query information. The outcome of interest is the product ranking. The model includes one constraint. Relevance scores are constrained to be monotonic in price. This constraint prevents situations where a firm can achieve a higher slot by increasing its prices, which can lead to positive own-price elasticities.

Aside from the adjustments to the relevance score and monotonicity in price, the estimation proceeds much like a standard learning to rank problem. I use normalized discounted cumulative gain (NDCG) as the loss function. I use cross-validation to select the optimal number of trees. The hyperparameters are shrinkage, interaction depth, and the out-of-bag fraction. I use .8 as a pilot out-of-bag fraction and Bayesian Optimization to select the shrinkage and interaction depth. Once I have the hyperparameters and number of trees, I fit the model on eight separate folds of the data. I estimate the model on distinct folds since the platform model relies on out-of-fold predictions. I then evaluate the fit of each model on a held out test data set.

Step 1 Results

Table 1.7.4 presents the out-of-sample performance for the first step of the platform model. Each row corresponds to a separate model, including a random benchmark, a model trained on the entire training data, each fold-specific model, and an ensemble from averaging the predictions of each fold-specific model. The columns correspond to loss functions; lower numbers mean better model performance. The two loss functions are NDCG, which is the loss function I used in model training, and concordant pair loss (Conc), which is the percent of pairwise pairs the model incorrectly ranks. Each fold-specific model has similar out-of-sample performance and performs well, correctly predicting rankings 72 percent of the time and performing better in predicting top-ranked products.[42] It is also worth pointing out that there is room for improvement, as the ensemble model (bottom row) performs notably better than each fold-specific model.

[42]The loss function for the model is the normalized discounted cumulative gain (NDCG), which can be challenging to interpret. I present the pairwise matching accuracy here since it is easier to understand.

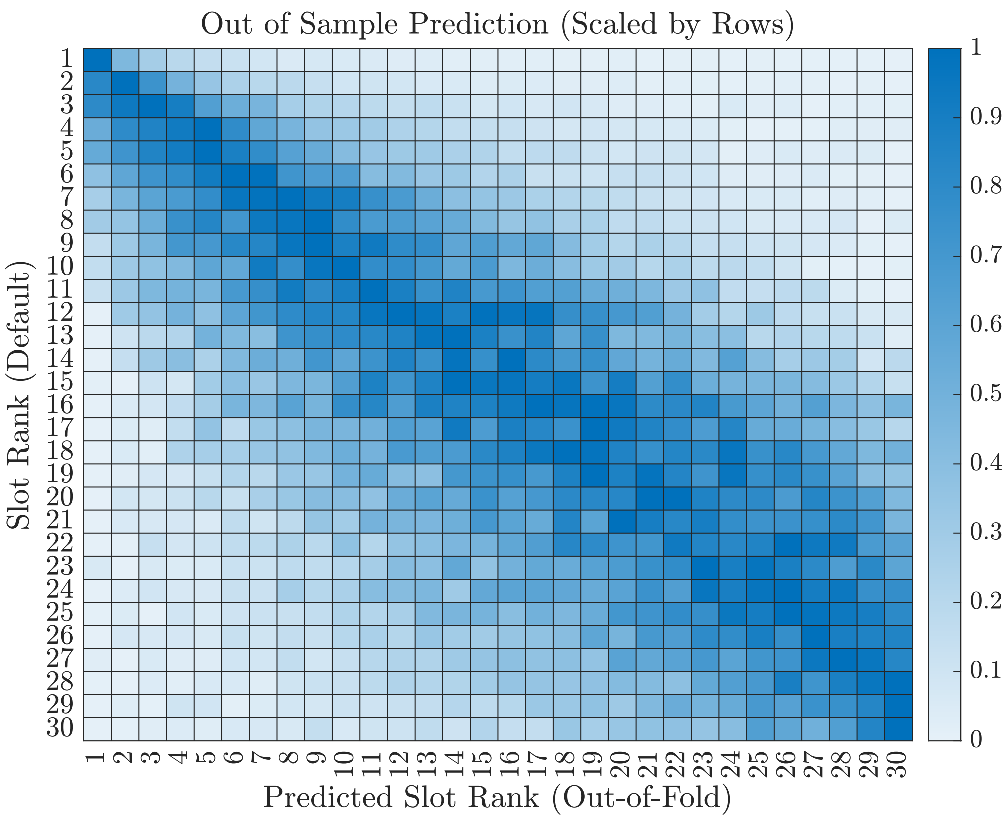
Table 1.7.4: Comparison of Model Results

Measure	Model	NDCG Loss	Conc Loss
1	Random Benchmark	0.175	0.506
2	LambdaMART (NDCG): Full	0.060	0.276
3	LambdaMART (NDCG): Fold 1	0.061	0.277
4	LambdaMART (NDCG): Fold 2	0.061	0.277
5	LambdaMART (NDCG): Fold 3	0.061	0.276
6	LambdaMART (NDCG): Fold 4	0.061	0.277
7	LambdaMART (NDCG): Fold 5	0.061	0.278
8	LambdaMART (NDCG): Fold 6	0.061	0.276
9	LambdaMART (NDCG): Fold 7	0.060	0.277
10	LambdaMART (NDCG): Fold 8	0.061	0.277
11	LambdaMART (NDCG): Ens	0.059	0.272

Note: NDCG is normalized discounted cumulative gain loss. Conc is concordant pair loss. Random benchmark uses random prediction. Ens (ensemble) averages the predictions from all eight folds.

Figure 1.7.2 visualizes the distribution of predicted rankings from the fold-specific models (treating their predictions as deterministic). The horizontal axis is the predicted slot, with 1 being the top-ranked product. The vertical axis is the observed slot. The dark diagonal means the predictions are in roughly the correct position. The darker region in the top left also illustrates that these models accurately predict the top products. This is important as most purchases and clicks occur in this top region of pages.

Figure 1.7.2: Distribution of Predicted vs Actual Slots



Note: Normalized distributions by row to adjust for different numbers of hotels appearing in search results

1.7.2.2 Platform Model Step 2: Sequential Logits

The first step model gives us the out-of-fold predicted deterministic component of relevance scores, $\hat{\psi}_{ijt}$, for each consumer-query-product, ijt.[43] The second step solved for the term that scales the deterministic portion of relevance scores, $\hat{\psi}_{ijt}$ to the scale of the random portion of relevance.

This second step estimates a slot-specific scale term, β_n^{slot}, on $\hat{\psi}_{ijt}$. For the first slot, this involves estimating a conditional logit with $\hat{\psi}_{ijt}$ as the right hand side variable, and an outcome of 1 if the target consumer-query-product is in the top slot. This regressions estimates the parameter β_n^{slot} from 1.7.24.

$$u^r_{ijtn}(\text{slot } n) = \beta_n^{slot}\hat{\psi}_{ijt} + \varepsilon_{ijt} \tag{1.7.24}$$

which give me the likelihood

$$P(j \text{ in slot } 1) = \frac{\exp(\beta_1^{slot}\hat{\psi}_{ijt})}{\sum_k \exp(\beta_1^{slot}\hat{\psi}_{ikt})} \tag{1.7.25}$$

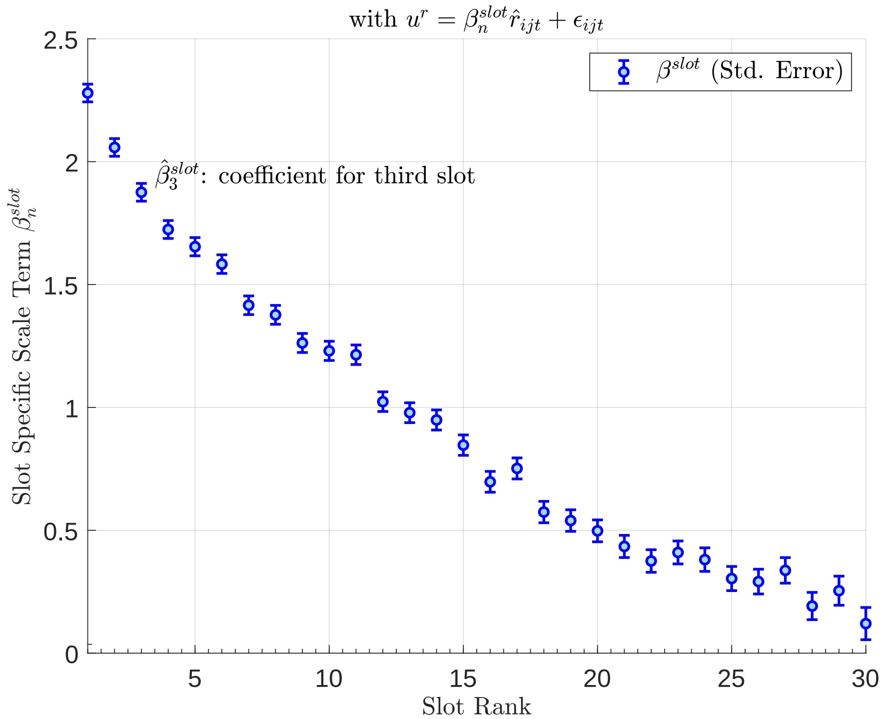
I then repeat this for target slot values of from 2 to 30. In each of these regressions, I use data for consumer-query-product with a slot at or below the target slot.

$$P(j \text{ in slot } n) = \frac{\exp(\beta_1^{slot}\hat{\psi}_{ijt})}{\sum_{k \notin \text{ slot 1 to } n-1} \exp(\beta_1^{slot}\hat{\psi}_{ikt})} \tag{1.7.26}$$

Figure 1.7.3 present the result from each of the conditional logit regressions. The parameters are higher for higher slots, meaning position on the page is more deterministic in $\hat{\psi}_{ijt}$ higher on the page.

[43] After step 1, I normalize the scores to be $N(0,1)$

Figure 1.7.3: Platform Model Sequential Logit Scale Parameters



In the supply side, I estimate own-price elasticities using the algorithms developed in step one, that predict $\hat{\psi}_{ijt}$, and the scale terms, β_n^{slot}, from step two.

1.7.3 Supply-Side Model: Hotel Pricing

The supply-side model captures hotel pricing behavior. In estimation, I use the seller's first-order conditions, the observed prices, and the expected quantities and elasticities from the demand and platform models to back out marginal costs. Since marginal costs depend on quantity, I use an IV approach with three-stage least squares to capture the relationship between marginal cost, quantity, and quantity squared.

Aggregation by Subperiod

The supply side (and counterfactual) requires some aggregation in terms of time of stay and time of search. In the hotel industry, prices change often. In part, these price changes serve as inter-temporal price discrimination. I define "subperiods" which are time of stay t and time of search t' pairs. In my primary specification, there are four subperiods per month, based on the combination of weekend vs weekday stays, and searches in advance of the stay or close to the date of the stay. This simplification allows for some inter-temporal price changes but provides enough hotel-subperiod observations to calculate the own-price elasticities necessary for the supply side and counterfactuals.

Seller's Subperiod Expected Profits

Restating the hotel's pricing problem, we have

$$\underset{p_{jtt'}}{\operatorname{argmax}} \mathbb{E}\left[(1-\varphi)\, p_{jtt'} - c_{jtt'} q_{jtt'} \,|\, \Omega_{jtt'}\right] \tag{1.7.27}$$

Seller foc

$$\frac{mc_{jtt'}}{1-\varphi} = p_{jtt'} + \left(\frac{\partial q_{jtt'}}{\partial p_{jtt'}}\right)^{-1} q_{jtt'} \tag{1.7.28}$$

were $p_{jtt'}$ is the price for room–night j, staying period t, and searching period t' is denoted by $p_{jtt'}$. To aggregate to the subperiod, I use the median observed $p_{jtt'}$. $mc_{jtt'}$ represents the opportunity cost of having the marginal unit available to sell in the next (sub)period.[44] $c_{jtt'}$ denotes average variable opportunity cost. The seller's information set, $\Omega_{jtt'}$, marks that sellers are aware of their own costs, the elasticity of demand, and the features and availability of competing products.[45]

The remaining elements on the right hand side of the seller foc (equation 1.7.28) are the expected quantity $q_{jtt'}$, and the inverted $\frac{\partial q_{jtt'}}{\partial p_{jtt'}}$, which depends on the own-price elasticity, the expected quantity, and the median price. These two elements depend on the demand and platform models.

I use the results from the demand model to estimate supply side observation weights. I then combine the results from the demand and platform models to calculate expected the expected quantities and own-price elasticity needed for the supply side model. I detail these procedures in Appendix A.6.1.

1.7.3.1 Marginal Cost Recovery and Specification

I back out estimated marginal costs, $\hat{mc}_{jtt'}$, using the seller's first-order condition in Equation 1.7.28, the observed prices, and the expected quantities and elasticities derived from the demand and platform models.

[44]It can also include additional expected profits that are conditional on the purchase, such as room service, dining, or gambling.

[45]These are the standard assumptions about the seller information with some extensions to account for the e-commerce environment.

Economies of Scale and Capacity Constraints

The supply-side model captures three key features of the accommodation industry: at low occupancy, hotels have economies of scale, and at high occupancy, hotels face increasing costs and capacity constraints. I capture these features by allowing the marginal cost to depend on quantity and quantity squared.

$$\hat{mc}_{jtt'} = mc^{\text{base}}_{jtt'} + \gamma_1 q_{jtt'} + \gamma_2 q^2_{jtt'} \tag{1.7.29}$$

Where $\hat{mc}_{jtt'}$ is the marginal cost estimate recovered from the hotel's first-order condition (FOC). A negative coefficient for $q_{jtt'}$ captures economies of scale, while a positive coefficient for $q^2_{jtt'}$ captures increasing costs at high occupancy and serves as a soft-capacity constraint.[46]

I cannot estimate Equation 1.7.29 directly since quantity depends on prices, which are decided endogenously. To address the endogeneity concern, I use an instrumental variable (IV) approach with BLP-type instruments Berry et al. (1995). The instruments are demand shifters, including features and availability of competing products in the same market and interaction terms of a hotel's own star rating with the distribution of star ratings in the market. With these instruments, I estimate the supply-side model via three-stage least squares.

First stage: IV for $q_{jtt'}$

The first stage instruments for quantity

$$q_{jtt'} = \alpha_1 x_{jtt'} + \alpha_2 z_{jtt'} + \varepsilon_{jtt'} \tag{1.7.30}$$

where $x_{jtt'}$ includes product features, and market-subperiod specific effects. The instruments,$z_{jtt'}$ include product features and availability of other products in same market, and own-star rating interactions.

[46] Farronato and Fradkin (2022) use a similar approach to modeling soft-capacity constraints in the hotel industry by estimating fixed and variable costs at low capacity, and increasing costs at 85% occupancy (a hockey stick-type function).

Second Stage: IV for $q^2_{jtt'}$

In the second stage, I instrument for $q^2_{jtt'}$ using the squared predicted values from the first stage, $\hat{q}^{\text{step 1}}_{jtt'}$.[47]

$$q^2_{jtt'} = \alpha_3 \left(\hat{q}^{\text{step 1}}_{jtt'}\right)^2 + \varepsilon_{jtt'} \tag{1.7.31}$$

Third Stage

In the third stage, I include the predicted values from the first stage, $\hat{q}^{\text{step 1}}_{jtt'}$, and second stage, $\hat{q}^{\text{step 2}}_{jtt'}$. The parameters of interest are γ_1 and γ_2.

$$\hat{mc}_{jtt'} = \hat{\beta} x_{jtt'} + \gamma_1 \hat{q}^{\text{step 1}}_{jtt'} + \gamma_2 \left(\hat{q}^2_{jtt'}\right)^{\text{step 2}} + \nu_{jtt'} \tag{1.7.32}$$

1.7.3.2 Supply-Side Results

The supply-side results, presented in table 1.7.5, are consistent with intuition: costs are higher for higher star-rating (tier) hotels and reflect the expected relationship with quantity, with decreasing costs at low quantity, characteristic of economies of scale, and increasing costs at high quantities, characteristic of increasing costs near capacity constraints. An additional specification includes star-specific γ's.

1.8 Personalized Recommendation Systems for Counterfactuals

Now, with a structural model of demand, platform product recommendations, and hotel pricing behavior, I turn to understanding the welfare effects of personalized recommendations. To do this, I first need to develop personalized recommendation systems. My model training approach is based on the winning entry in the Expedia Personalization competition. I use an ensemble of LambdaMARTs, which are learning-to-rank algorithms that use gradient-boosted decision trees, with NDCG as the loss function.

1.8.1 Training Four Recommendation Systems

As with estimating demand, a challenge in training recommendation systems is that slots influence consumer choices, but slots are highly correlated with product features. This is

[47] Skipping this step, and including directly including $\hat{q}^{\text{step 1}}_{jtt'}$ in the regression on marginal cost would not produce the same results, and would be a "forbidden regression". For more info see chapter 9 of Wooldridge (2010)

Table 1.7.5: IV Regression Analysis Results

Variable	**Estimate**
(Intercept)	-0.373
	(0.503)
$\hat{q}_j^{(1)}1$	-0.195***
	(0.037)
$\hat{q^2}_j^{(2)}$	0.032***
	(0.008)
Two/Three-Star	0.577***
	(0.088)
Four-Star	1.010***
	(0.086)
Five-Star	2.688***
	(0.111)
Product Feature Controls: ✓	
Location Desirability Controls: ✓	
Month–Weekend–Subgroup Controls: ✓	
Observations: 3492	
Degrees of freedom: 3437	
RMSE: 0.748	
R^2: 0.656, Adjusted R^2: 0.651	
F-statistic: 121, p-value: 0.00	
First-stage F-Statistic: 103	

Note: Instruments for $\hat{q}_j^{(1)}$: Mean product features of competing products, availability of other products, and own star rating interacted with the distribution of star ratings of other products.

where the randomized data are incredibly useful. I train these recommendation systems using the subset of Expedia data where they randomized the product rankings.[48] I use the data from the five top markets, as there could be information spillovers from one market to another. For example, preferences may be similarly correlated across markets.

For the outcome variable, relevance score, I follow the approach from the original competition rules, where relevance scores are 5 for bookings, 1 for clicks, and 0 for impressions.

I train four increasingly personalized versions of the recommendation system. In counterfactuals, this helps understand not just what would happen with the most personalized recommendation system (possible with these data) but how welfare would change as we increase personalization from the least personalized to the most personalized. I adjust the level of personalization based on the variables available to the algorithms. The least personalized recommender only uses data on products. The next includes additional data on the consumer

[48] In fact, the winning entry from the Kaggle competition only used random data to train their models.

queries. This is information actively volunteered by consumers, such as length of stay and if they are traveling with children. The next recommendation system uses personal data based on the consumer's location, distance to the destination, booking window, and time of search. The most personalized recommendation system includes data on each consumer's past purchases, such as the average price, star rating of their previous purchases, and other tracked information.

Common Recommendations: Product features, price and availability on competing OTAs.

Query Adjusted: Added query features (e.g., length of stay, number of nights).

Personalize: Added consumer observables (e.g., booking window, consumer country).

Most Personalized: Included past transactions, tracked navigation data.

I use an ensemble approach for each recommendation system, in which I train multiple versions of LambdaMART and take the average of their predictions. Each of the four recommendation systems consists of 170 underlying LambdaMARTs. I constrain each of the 170x4 models to be monotonic in price.

1.8.2 Validating Recommendation Systems

I validate the models by estimating out-of-sample performance in predicting purchases and clicks. I should find that models with access to more personalized data have better out-of-sample performance. Table 1.8.1 presents the results. These models match that pattern with out-of-sample model performance increasing with personalization.

Table 1.8.1: Out of Sample Comparison of Model Results

Measure	Model	NDCG Loss	Conc Loss	MAP	MRR
1	Random Benchmark	0.673	0.480	0.850	0.846
2	LambdaMART (Ensemble): Base Info	0.544	0.302	0.699	0.692
3	LambdaMART (Ensemble): with Query Info	0.540	0.301	0.695	0.686
4	LambdaMART (Ensemble): Personalized Basic	0.537	0.299	0.692	0.681
5	LambdaMART (Ensemble): Personalized Full	0.533	0.300	0.686	0.676

Note: NDCG is normalized discounted cumulative gain loss. Conc is concordant pair loss. MAP is mean average precision, MRR is mean reciprocal Rank

1.9 Counterfactual Simulations

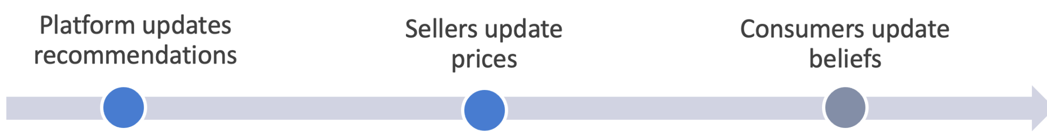
This section presents the counterfactual simulations. I use the structural model to evaluate the welfare effects of deploying personalized recommendation systems.

1.9.1 Counterfactual Setup

In the counterfactuals, I make a few necessary simplifications. In the hotel industry, prices change often. As a baseline counterfactual, I impose a sub-period uniform pricing constraint. In this setup, hotels can set four unique prices for room–nights in each month along two dimensions: weekends versus weekdays and searches in advance of the stay or close to the stay date. These subperiods match the supply side of the model. This simplification allows for some inter-temporal price changes but provides enough hotel–subperiod observations to calculate the own-price elasticities necessary for the supply side.

Next, I solve for the counterfactual equilibrium using the baseline, and the four recommendation systems from section 1.8. For each of these recommendation systems, I solve the equilibrium in three distinct phases: First, the platform updates the recommendation system. Second, sellers update prices. In the next phase of the project will include a third step where consumers update their beliefs about the recommendation system.[49]

Figure 1.9.1: Counterfactual Timing



My outcomes of interest are seller profits, quantity sold, platform revenue, and consumer surplus. I repeat the counterfactual analysis under different supply-side assumptions, fixed marginal cost, common economies of scale and soft capacity constraints, and star-level economics of scale and soft capacity constraints. There are 60 counterfactuals based on five recommendation systems, three supply-side assumptions, and updating versus not updating prices.

1.9.2 Solving for Equilibrium

Here I briefly describe the process to solve for the new equilibrium. I first solve for the baseline counterfactual of subperiod uniform pricing, using the platform model to generate

[49]This is a work in progress, these results will be available shortly.

recommendations, and a contraction mapping of the firms first order conditions. The resulting prices serve as the baseline for the remaining counterfactual.

Without Price Updates. To solve for the new equilibrium without price updates, I take the structural model and replace the product recommendation system that provides the deterministic portion of relevance scores with the target recommendation system ψ. Then keeping prices fixed, I calculate quantities, gross booking revenue, and firm profits, and consumer surplus. The consumer surplus depends on the utility of the predicted purchases, and search costs of the predicted clicks.

With Price Updates. To solve for the new equilibrium with price updates follows the same process as without price updates, except I solve for new prices with a contraction mapping of the hotel's first order conditions.

Counterfactual Limitations. As noted above, one limitation of the counterfactuals comes from the need to aggregate to the subperiod level. There are a few other limitations, as this is a work in progress. Right now the counterfactual focus on four subperiods for the top market. There are over 700 unique hotels in the top market, but many appear rarely. I hold the prices of hotels that appear less than five times in a given subperiod fixed.

1.9.3 Counterfactual Results

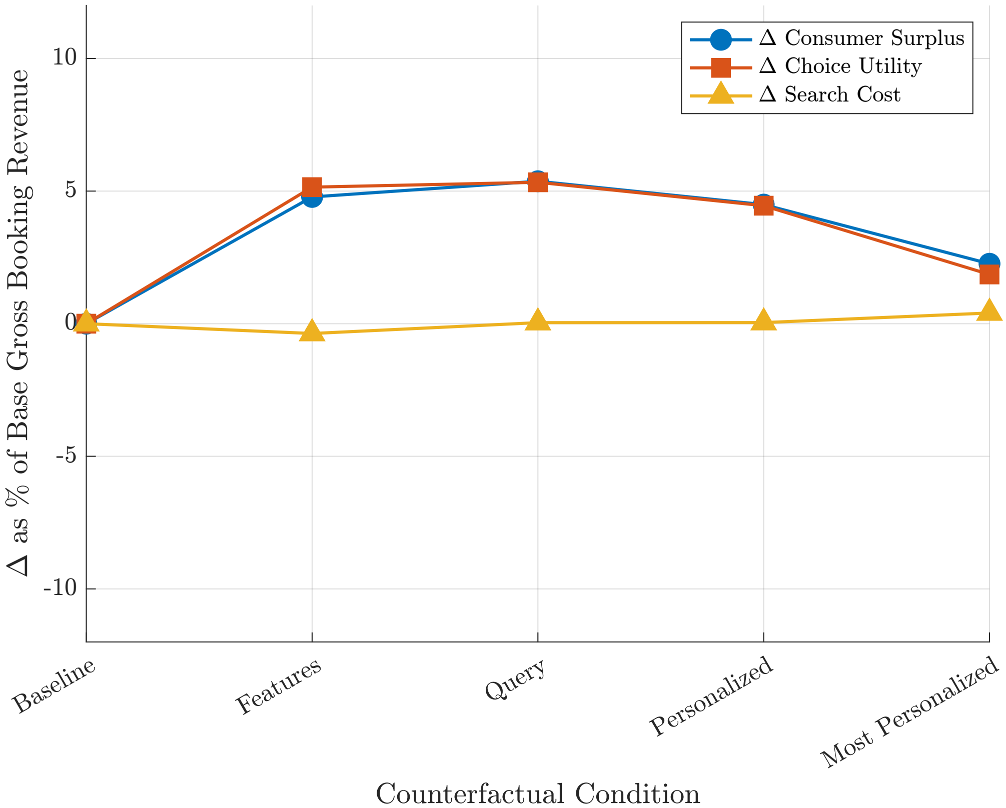
Here I present the counterfactual results for the primary specifications. In the primary specifications, the supply side includes star-rating specific marginal cost functions. The extended set of results are included in Appendix A.7.

1.9.3.1 Counterfactual without Price Updates: Welfare Gain

Holding prices fixed, I find a welfare gain from personalized recommendations. Figure 1.9.2 plots the consumer surplus results, including the utility of final choices and search cost of clicks. Table 1.9.1 includes the complete set of outcomes, quantity, gross booking revenue, and hotel profits. It also includes an approximate platform revenue, assuming they take a 10% commission. I present the consumer welfare numbers relative to the baseline since, with discrete choice models, we can identify differences in consumer surplus but not the absolute level.

I do note an marginal decrease in consumer surplus going from the query level to personalized and most personalized. If we were to take the demand model as the truth, this

Figure 1.9.2: Counterfactual without price updates and with star-level economies of scale and soft capacity constraints: *welfare gain*



Note: Change in values represented as a percent of baseline gross booking revenue.

would be due to potential over-fitting of the personalized recommendation systems. However, I suspect the more likely explanation is a limitation of the current demand specification. The personalized and most personalized models were trained using variables that the demand model does not include. As a next step, there are two options to address this concern: 1) A new demand model with more parameters to capture heterogeneous preferences. 2) Using a conditioning on individual tastes (COIT) post estimation procedure to make personalized welfare predictions (Revelt and Train, 2000).

The welfare gains come from consumers choosing higher utility products and lower incurred search costs. Gross booking revenue remains relatively unchanged, indicating that consumers choose higher utility products but do not, on average, substitute from the outside option to one of the inside goods.

Back of the Envelope Welfare Change The gain in consumer surplus is 2.3% of total booking revenue. If we want an idea of the scale of the welfare effects of going from baseline

Table 1.9.1: Counterfactuals with No Price Updates, with Star-Level Economies of Scale and Soft-Capacity Constraints

		Recommendation System			
Outcomes	**Baseline**	**Features**	**Query**	**Personalized**	**Most Personalized**
Quantity	508.5	505.9	505.8	505.7	505.8
Gross Booking Revenue ($100s)	1,809.37	1,804.72	1,806.78	1,806.16	1,807.52
Hotel Profits ($100s)	984.96	985.14	985.18	984.75	984.99
Approx Platform Revenue ($100s)	180.94	180.47	180.68	180.62	180.75
Consumer Welfare					
Δ Consumer Surplus ($100s)	0	86.48	97.18	81.25	40.89
Δ Choice Utility ($100s)	0	93.17	96.47	80.50	33.59
Δ Search Cost ($100s)	0	-6.69	0.71	0.75	7.30

to most personalized, we can scale the consumer surplus change by Expedia's gross booking revenue for the same year, 2013. This calculation would imply ~$0.9 billion increase in consumer surplus. These results are consistent with previous literature that finds welfare gains from improving recommendation systems while holding prices fixed.

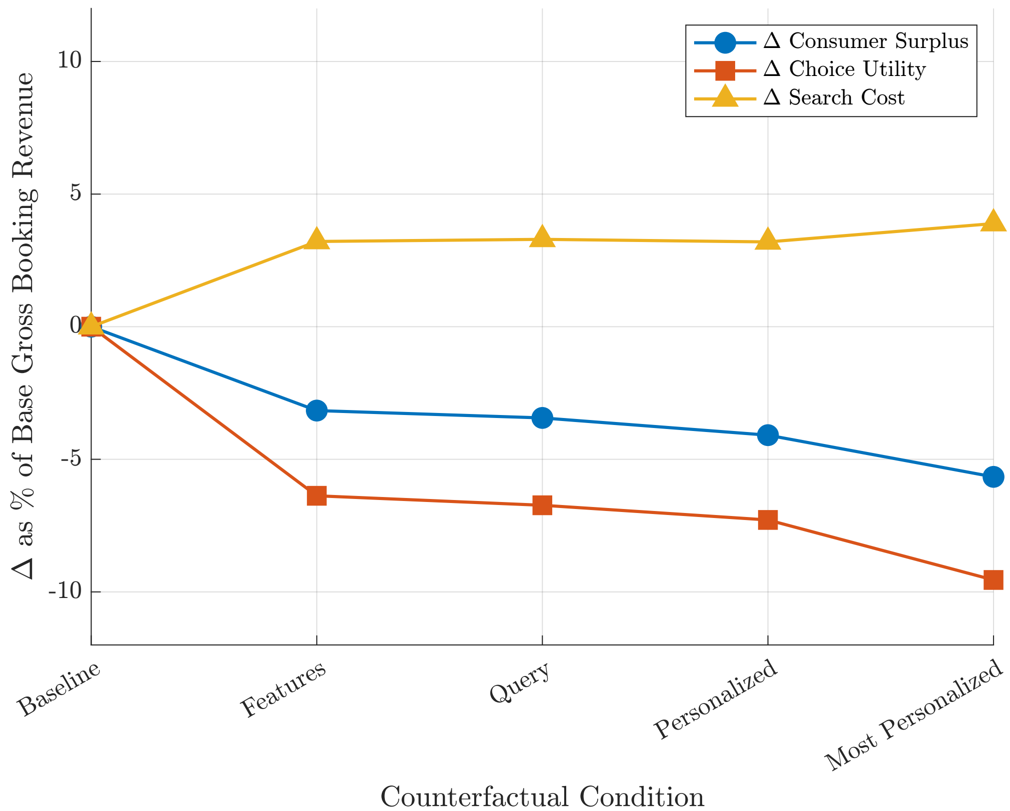
1.9.3.2 Counterfactual with Price Updates: Welfare Loss

Once sellers can update prices, I find a welfare loss from personalized recommendations. Figure 1.9.3 plots the consumer surplus results, including the utility of final choices and search cost of clicks. Table 1.9.2 includes the complete set of outcomes.

The results are consistent with a market becoming less competitive. Going from baseline to most personalized, sellers increase prices and see a 4.9% increase in profits. I also find a 4.5% decrease in quantity. Consumer surplus decreases by 5% of baseline gross booking revenue. Scaling these results up by the gross booking revenue of Expedia in 2013 would correspond to a ~$2 billion decrease in consumer surplus. This would represent a total welfare loss, as the decrease in consumer surplus is nearly double the increase in hotel profits.

From this counterfactual simulation, I find that ignoring seller price adjustments causes considerable differences in the estimated impact of personalization. Without price adjustments, personalization would increase consumer surplus. However, sellers have an incentive to increase prices. Indeed, I find a welfare loss from personalization.

Figure 1.9.3: Counterfactuals with price updates and star-level economies of scale and soft capacity constraints: *welfare loss*



Note: Change in values represented as a percent of baseline gross booking revenue.

Table 1.9.2: Counterfactuals with Star-Level Economies of Scale and Soft-Capacity Constraints

		Recommendation System			
Outcomes	**Baseline**	**Features**	**Query**	**Personalized**	**Most Personalized**
Quantity	517.6	495.2	494.8	494.2	494.3
Gross Booking Revenue ($100s)	1,830.09	1,825.62	1,829.00	1,827.90	1,829.79
Hotel Profits ($100s)	974.02	1,020.00	1,021.20	1,021.32	1,022.03
Approx Platform Revenue ($100s)	183.01	182.56	182.90	182.79	182.98
Consumer Welfare					
Δ Consumer Surplus ($100s)	0	-27.37	-62.97	-66.19	-92.02
Δ Choice Utility ($100s)	0	-75.16	-124.19	-118.06	-158.50
Δ Search Cost ($100s)	0	47.79	61.22	51.88	66.48

1.10 Conclusion

In this paper, I explore the welfare effects of personalized recommendations in digital markets using data from Expedia Group. While this paper focuses on Expedia, it addresses a familiar dynamic between sellers and e-commerce platforms in the increasingly digital economy. The platform chooses its platform design, including the recommendation system, but third-party sellers, in this case hotels, set prices. Personalized recommendations can improve consumer welfare through the long-tail effect, where consumers find products that better match their tastes. However, third-party sellers, facing demand from better-matched consumers, may be incentivized to increase prices.

I develop a structural model of demand, platform product recommendations, and hotel pricing behavior to quantify the tradeoff between match quality and price competition. On the demand side, this paper proposes an optimal sequential search model where consumers have beliefs about the joint distribution of product features and recommendations, form consideration sets through clicks, and make a final purchase decision from their consideration set. For the product recommendation model, I use a "model of a model" machine learning approach to reverse engineer Expedia's default recommendation system. Combining the results from the demand and recommendation system models allows for the supply-side model where capacity-constrained hotels consider how changes in price impact position on the page in search results.

In addition to the structural model of demand, platform recommendations, and seller pricing behavior, I develop four increasingly personalized recommendation systems. I use an ensemble of LambdaMARTs, a popular machine-learning algorithm for ranking problems. I use Expedia data to train the recommendation systems, where Expedia randomized product rankings.

In my counterfactuals, I find that ignoring seller price adjustments would cause considerable differences in the estimated impact of personalization. Without price adjustments, personalized recommendations would increase consumer surplus by 2.3% of total booking revenue ($0.9 billion). However, once sellers update prices, personalization would lead to a welfare loss, with consumer surplus decreasing by 5% of booking revenue ($2 billion). This paper provides actionable insights relevant to researchers, platforms, and policymakers and highlights an overlooked concern in e-commerce platform research and regulation: Better recommendation systems may reduce competition and harm consumer welfare. This finding is important to consider as e-commerce platforms' access to personal data grows, and technological improvements allow platforms to deploy increasingly sophisticated recommendation systems.

Policies that mitigate these pricing effects could be available to regulators and platforms. In the next step of this project, I plan to consider a policy alternative that tunes the recommendation systems to account for price competition. Operationally, this would involve increasing or decreasing the price sensitivity of the personalized recommendations to achieve a policy goal, for example, maximizing total surplus or equating equilibrium prices with those that would prevail without personalization. These price-tuned recommendation systems can potentially improve platform revenue, hotel profits, and consumer surplus.

CHAPTER 2

Differential Impacts of Online Ratings in the Market for Medical Services

with Sonal Vats and Michael Luca

2.1 Introduction

Credence goods are a type of good with qualities that are difficult or impossible to fully judge by a consumer even after purchase and consumption (Darby and Karni, 1973). Medical services are a prominent example of credence goods. In the market for physicians services, consumers face both ex-ante and ex-post uncertainty regarding the quality of care.

In the market for medical services, consumers face a challenge in both selecting a physician and also judging the quality of the services received from that physician. In the past, consumers have relied heavily on social learning to resolve these asymmetries. For example, consumers may ask their peers to recommend a physician. In fact, the National Institute for Aging tells patients to "ask people you trust" for physician recommendations.

2.1.1 The Role of Online Ratings

Online ratings are an increasingly important driver of economic activity and consumer decision-making. The three industries where online ratings are most viewed are restaurants, hotels, and healthcare.[1] Through websites like ZocDoc, a unique website that integrates physician profiles, patient reviews, and appointment scheduling, online reputation, is a potentially important source of information for consumers about physicians. The large-scale distribution of information from numerous other consumers, could help resolve information asymmetries among a much broader peer group than was previously been possible.

[1] https://www.brightlocal.com/research/local-consumer-review-survey-2020/

On the provider side, these subjective reviews could be a source of feedback to physicians they would not receive otherwise. In health care markets, there is strong evidence that public disclosure of quality data has been effective in better matching patients with products and providers. Studies find that consumers tend to prefer higher-quality providers. Dafny and Dranove (2008) and Jin and Sorensen (2006) find that the publication of report cards boosted the market share of insurance plans that received higher scores. Another example is Bundorf et al. (2009), who find that fertility clinics with high birth rates gained market share after the US Center for Disease Control and Prevention began publishing success rates in 1997. Using a fixed effects framework, Wang et al. (2011) find that surgeons who received poor ratings on Pennsylvania CABG report cards experience a decrease in patient volume. Similarly, Cutler et al. (2004), find that lower ranking hospitals in New York lost market share, especially among less severely ill patients. However, disclosure can also harm consumers if sellers can superficially boost their performance through window dressing. For example, Dranove et al. (2003) find that cardiac surgery report cards in New York and Pennsylvania led to selection by providers, which suggests a serious downside risk on quality reporting in health care. Werner et al. (2009) and Feng Lu (2012) find similar evidence with the Nursing Home Quality Initiative.

It is unclear whether consumer review websites should significantly affect markets for credence goods. Consumer review websites help fill the void left by the absence of any government or nonprofit agency assuming the role of information provider on primary care physician quality. Consumer reviews can also be a complement or substitute for existing information—education, board certification, and malpractice claims—on physicians, some of which may not be easily available or understood by a lay person. Alternatively, a consumer writing a review cannot fully evaluate the treatment or service received, since they are unfamiliar with the intricacies of the medical knowledge possessed by the primary care physician. Further, patient-created reviews can be difficult to interpret—they reflect the views of a non-representative sample of patients and are subjective.

Notwithstanding these challenges, an emerging body of economics literature studies the impact of online ratings on demand for medical services. McCarthy et al. (2022) combine Yelp reviews with claims data to show that patients are willing to travel further to receive care from hospitals with higher Yelp ratings. Brown et al. (2023) study demand for General Practitioner (GP) offices in England and show that patient demand from low-income neighborhoods responds sharply to summary star ratings. Collectively, these studies underscore that online ratings could be a driving force in healthcare decision-making.

There is also reason to suspect that the returns to online reputation are different for different physicians. In a field study, Chan (2023), finds that signals of doctor quality reduce

90% of the racial gaps in willingness to pay for doctors. Brown et al. (2023) find that the impact of ratings could be mediated by private information. Other work shows evidence of the differential impact from other forms of reputation, for example Sarsons (Sarsons) evaluates how patient deaths impact referrals to surgeons and finds that female surgeons experience a larger drop in referrals after a patient death.

Zocdoc presents an ideal context to study the impact of online reviews as it has the following notable features: 1) A discontinuous rating system; 2) Consumers face little variation in prices across providers in their insurance network; And 3) A closed-loop review system. Zocdoc displays ratings on a scale of one to five stars, with overall average ratings, rounded to the nearest half-star. We take advantage of the fact that Zocdoc rounds ratings to the nearest "half-star." As a result, two physicians with nearly identical ratings can straddle the cutoff to display 4.5 versus 5-stars. These may be viewed as very different by consumers, even if the underlying quality is quite similar. We use this natural experiment to estimate the causal impact of ratings on booking volumes using a regression discontinuity design. Further, we explore the differential impacts of rating by repeating the analysis for economically interesting subgroups of physicians.

2.2 Background on ZocDoc.com

Launched in 2007, Zocdoc is an online medical care search, and scheduling service. Freely available to patients, the website enables patients to search for physicians by insurance, location, specialty, procedure, hospital affiliation, gender, and languages spoken. Based on the selection criteria ZocDoc provides patients with a list of physicians, patients can view open slots in physicians' schedules and make an appointment online. According to ZocDoc, most of the appointments happen in 24-72-hour window.[2] Zocdoc appointment service was initially limited to dentists in Manhattan, as of 2013, Zocdoc claims to serves 40 percent of the U.S. population across more than 1,800 cities. More than 2.5 million patients use Zocdoc to find doctors every month.[3]

A patient looking for physicians on Zocdoc can use the website's search feature, in Figure 2.2.1, to search for physicians based on specialty, location, and insurance. For example, Figure 2.2.2 shows a snapshot of the list of physicians in Boston, with no restriction on the type of insurance they accept. This search results page presents the patient with an ordered list of physicians, displaying displaying the physician's photograph, practice address, rounded average rating, main specialty, medical degree, hospital affiliation, and all open slots in their

[2] 12 Facts about Zocdoc Users

[3] Zocdoc Announces Patients Have Booked More Than 1,000 Different Procedures Through Its Free Service

appointment schedule for the current week. Clicking on any physician name takes the patient to the physician's profile page which displays additional information about the physician's education, specialty, languages spoken, types of insurance accepted, and also displays the detailed ratings and text reviews left by any patient. Figure 2.2.3 display the physician profile and the individual patient ratings as they appear on Zocdoc's physician profile page.

Figure 2.2.1: Search Page

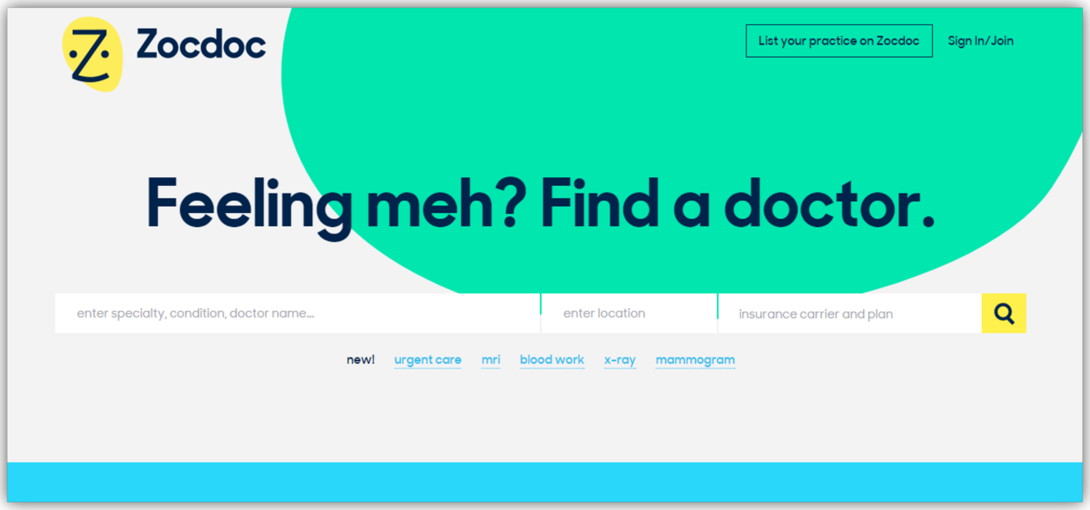


Figure 2.2.2: Example Zocdoc Search Results

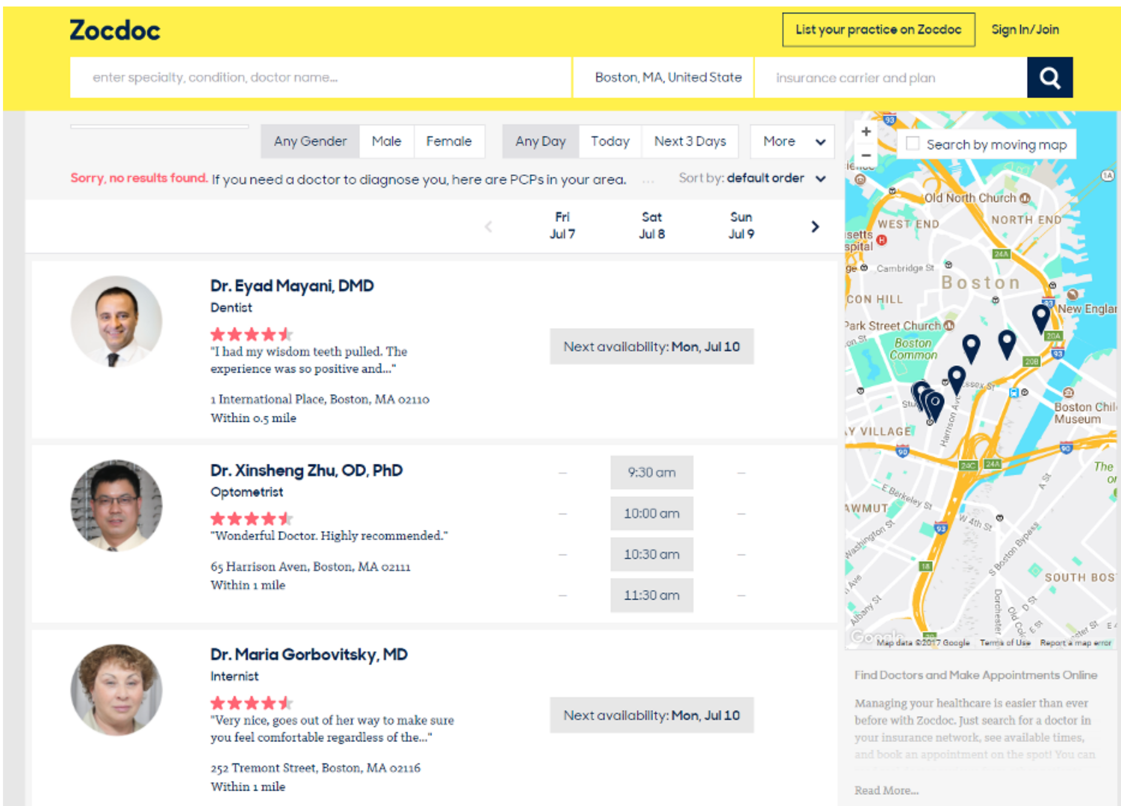
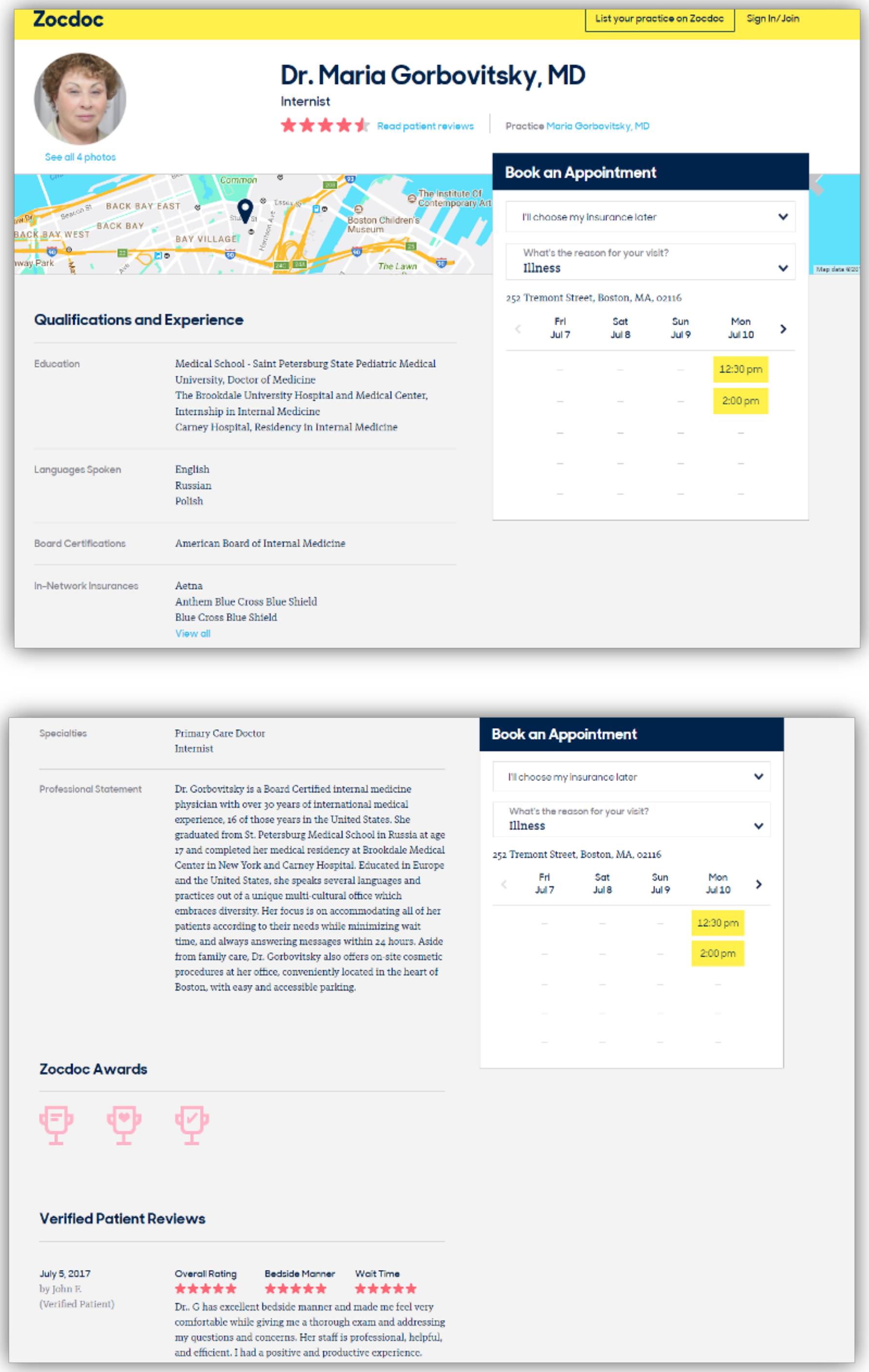
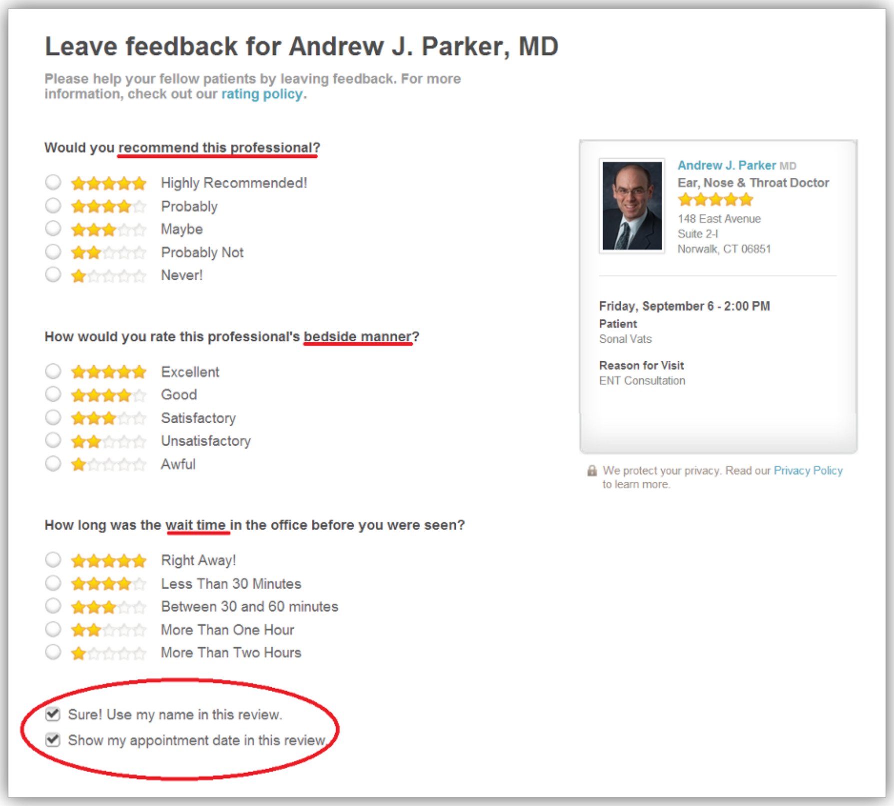


Figure 2.2.3: Example Profile Page



After an appointment, Zocdoc emails a thank-you note, encouraging patients to review and rate (from 1-5 stars) their physicians for bedside manner, wait time and overall impression, as illustrated in Figure 2.2.4. The patient can then rate the physician they visited and can also enter a text review. Once a review is written, anyone (with or without an account) can access the website and read the review. Patients will come across reviews within the context of the search for a physician. This allows the patients, looking for physicians on Zocdoc, to compare and assess them on common quality characteristics. Since each verified patient is encouraged to leave a review, it may not be that patients who have had extreme experiences, and who are proactive, are the only ones to leave reviews.

Figure 2.2.4: Post Appointment Feedback Prompt



During the time of our study, Zocdoc was free to patients and followed a subscription model for providers. Physicians can choose to subscribe by paying a monthly subscription of $250. For each subscribing physician the website has a profile page with 'verified' credentials, and patient–submitted reviews. Subscribing physicians could benefit by attracting new patients, and by filling the last-minute cancellations and postponements (10-20 percent of total appointments) by Zocdoc patients. However more recently, Zocdoc proposed pricing changes have been a source of concern for physicians.[4]

2.3 Data Construction

We construct a novel dataset using Zocdoc. These data include physician professional information, patient generated reviews, and appointment schedules. We focus on primary care physicians in eight metropolitan divisions where Zocdoc has a significant presence, with New York City being by far the largest. Our data includes over 5.8 million offered appointments from over two thousand primary care physicians, with 94% of profiles having at least eight reviews. Table 2.3 has the number of appointments and physicians per metropolitan division:

Table 2.3.1: Offered Appointments and PCPs by Metro Division

Metro Division	**Appts.**	**PCPs**
Boston, MA	86,512	117
Cambridge-Newton-Framingham, MA	71,159	68
Chicago-Naperville-Arlington Heights, IL	795,000	331
Fort Lauderdale-Pompano Beach-Deerfield, FL	184,185	80
New York-Jersey City-White Plains, NY-NJ	3,629,392	1,291
San Francisco-Redwood City-South San Francisco, CA	69,384	31
Silver Spring-Frederick-Rockville, MD	232,236	82
Washington-Arlington-Alexandria, DC-VA-MD	774,756	305

The data collection process works in three steps. In the first step, it uses Zocdoc's search engine, to compile a list of Primary Care Physicians and query specific search ranks. The second step collects profile information for each physician in the list. The third step collects all of each physician's reviews and available appointments for the next 35 days. The first two steps were repeated monthly. The third step was repeated daily at 2am. This exercise was repeated for a period of over one year, starting February 22, 2016 and running through

[4] CNBC: Zocdoc Price Surge has Doctors Fretting

April 17, 2017, and ended after updates to Zocdocs's website starting in March 2017 make our data less reliable.

For the first step, each search would yield a maximum of 50 results, so it was necessary to repeat each search with narrow filters in order to construct a comprehensive list of physicians in the area. In this step the search was repeated for each zip code, physician type, appointment type, and physician gender. Along with a list of physicians, we also collected a query specific search rank. The query specific search rank tells us the page rank of each physician for a given search. The first step is repeated once a month to search for new physicians.

For steps 2 and 3 we collect the following information each month:

Each month we also collect physician profile pictures. We then use Microsoft's Face API to collect additional demographic information.

Table 2.3.2: Data Collection of Physician Profiles, Ratings, Reviews, Practice Information, and Availability

Category	Variable	Visibility
Physician Profile Information (Collected Monthly)		
	Profile URL	html
	Doctor Name	Landing Page
	Title	Landing Page
	Specialty: Primary Care, Internist	Landing Page
	Badge - Rapid registration	Profile Page
	Badge - See You Again	Profile Page
	Badge - Speedy Response	Profile Page
	Badge - Scheduling Hero	Profile Page
	Practice Name	Profile Page
	Specialties	Profile Page
	Education: Medical School and Residency	Profile Page
	Hospital Affiliations	Profile Page
	Languages	Profile Page
	Board Certifications	Profile Page
	Awards and Publications	Profile Page
	In Network Insurances	Profile Page
	Doctor Code	html
	Professional Statement	Profile Page
Ratings and Reviews (Collected Daily)		
	Rounded Rating Overall Rating (Half-Star)	Landing Page
	Review level - Patient Name	Profile Page
	Review level - Date	Profile Page
	Review level - Overall Rating (1-5)	Profile Page
	Review level - Wait Time Rating (1-5)	Profile Page
	Review level - Bedside Manner Rating (1-5)	Profile Page
	Review Text	Profile Page
Practice Information (Collected Daily)		
	Address	Profile Page
	Coordinates: Latitude and Longitude	html
	Location ID	html
Schedule and Appointments (Collected Daily)		
	Available Illness Appointments	Profile Page*
	Available New Patient Appointments (next 35 days)	Profile Page*
	Appointments Start Times	Profile Page*
	Appointment Locations	Profile Page*
Search Information (Recorded Monthly)		
	Gender	Search Filter
	Zip Code	Search Filter
	Page Rank	Landing Page

* indicates 3-5 days of availability on the landing page, additional availability on profile and scheduling pages

2.3.1 Platform Updates

Our collected data spans from 2016 to 2017, a timeframe that proves particularly fitting for our study. The platform has since introduced numerous updates, adding complexity and making a similar study considerably more challenging for researchers today. Notably, it now displays an average rating rounded to the nearest hundredth, a departure from its previous practice of rounding off the overall rating to the nearest half-star, the platform design feature on which our natural experiment depends. The platform's redesigned landing page and profile pages provide users with an array of additional information and navigation options, which, while enhancing the user experience, obscures the simplicity of user interactions that our study aims to examine.

2.4 Empirical Strategy

Zocdoc employs a system that prominently displays on the landing page a physician's average overall rating rounded to the nearest half-star, with detailed rating information relegated to the profile pages. For example, as illustrated in Figure 2.4.1 a physician with a 4.74 average rating will be rounded down to 4.5 stars, while a physician with 4.75 stars will be rounded up to 5 stars. This allows us to examine observations with nearly identical underlying average ratings but a half star difference in the rating displayed to consumers. We leverage this rating system to analyze the impact of ratings on patient volumes using a regression discontinuity design. We supplement this with comprehensive profile data to evaluate the differential impacts of ratings by gender, number of ratings, and hospital affiliation.

2.4.1 Discontinuous Rating System

We specifically use a regression discontinuity design with multiple cumulative cutoffs, this exploits the fact that the rating system displays the rounded half-star rating on the landing page. We use the multiple cumulative cutoff approach since it has the interpretation that there is a different cutoff for each half star and that the treatment is different at each cutoff. For example, at the 4.25 cutoff, the control is 4-stars and the treatment is 4.5-stars. At the 4.75 cutoff, the control is 4.5-stars and the treatment is 5-stars.

For our preliminary analysis, we mainly focus on the 4.75 cutoff since that has the most observations in our data. Figure 2.4.2 plots the distribution of ratings by appointments offered in our primary sample. The data are left skewed with most appointments belonging to physicians with at least a 4.5 star rating.

Figure 2.4.1: Rounding of Overall Ratings

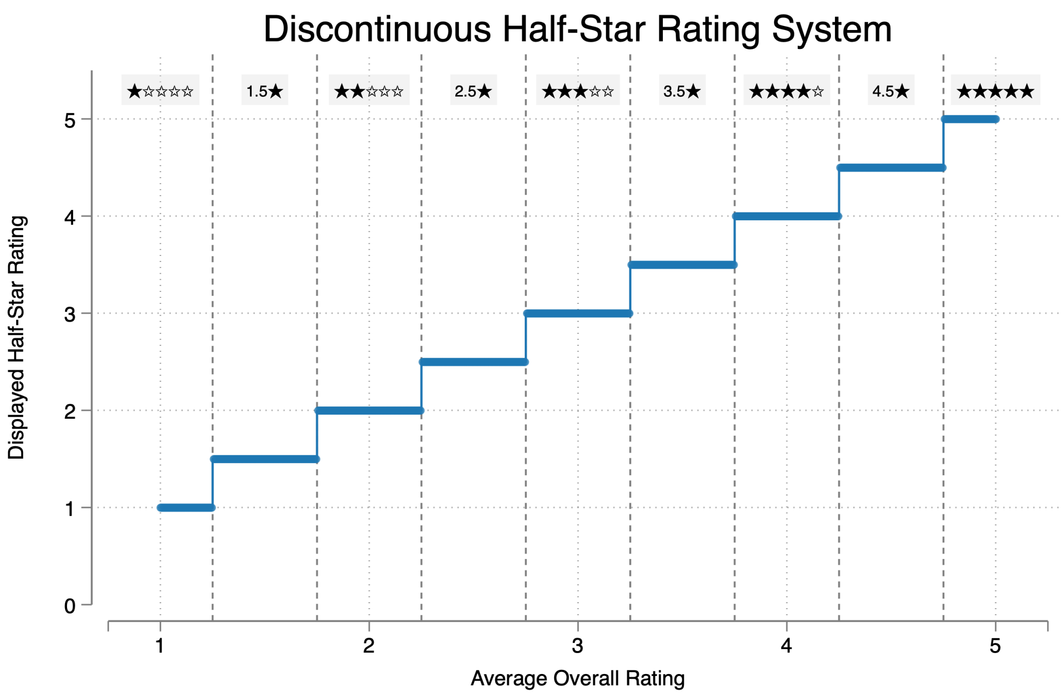
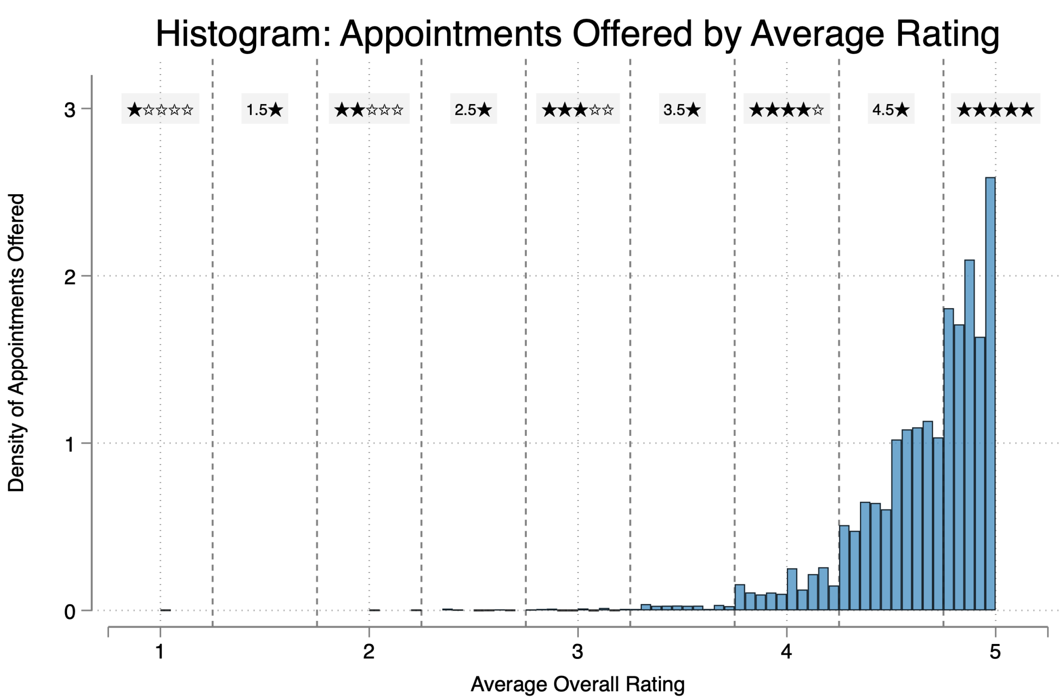


Figure 2.4.2: Appointments by Average Rating



The cutoffs faced by a physician on a given day are a deterministic function of the physician's average overall rating. The treatments at each cutoff are different in some respects. We formalize the notation for the average rating, X_j, the displayed star rating, s_j, and the relevant cutoff values, c_i.

- average rating: $X_i \in [1, 5]$
- displayed star rating: $s_i \in \{1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5\}$
- cutoff values: $c_i \in \{1.25, 1.75, 2.25, 2.75, 3.25, 3.75, 4.25, 4.75\}$

Next, we can write τ_i, the local causal treatment effect of an increase in a half-star. With the treatment is local to the average rating X_i exactly equal to cutoff c_i:

$$\tau_i = E[Y_j(s_i) - Y_j(s_{i-0.5})|X_j = c_i] = \lim_{x \downarrow c_i} E[Y|X_j = x] - \lim_{x \uparrow c_i} E[Y|X_j = x]$$

Cattaneo et al. (2020) point out that, in situations with multiple cumulative cutoffs, we can use each observation to estimate two different treatment effects. For example, a physician with an average rating of $X_i = 4.2$ could be used to estimate treatment effects for $\tau_{i=4}$ and $\tau_{i=4.5}$. In this case, the physician with an average rating of $X_i = 4.2$ would be considered treated when estimating $\tau_{i=4}$ and control when estimating $\tau_{i=4.5}$.

We estimate each τ_i cutoff-by-cutoff with local polynomial methods, and asymmetric data-driven bandwidth selection using the `rdrobust` package. For our primary specification, we use a first-order local polynomial with a triangular kernel function to construct the point estimator τ_i. We use a second-degree local polynomial for bias correction. To account for the panel structure of our data, we estimate cluster robust nearest neighbor variances.

2.4.2 Primary Outcomes

Our primary outcome of interest is weekly patient volumes. We identify offered appointments as booked if they disappear from schedules up to three days before the date of the appointment. As a pilot we chose three days to avoid measurement error as many physicians choose not to list appointment same day or two days before the appointment date. These estimates should inform both: 1) how consumers use ratings to select primary care physicians, and 2) the gains to physician for having a better online reputation. For this draft we use the inverse hyperbolic sign (IHS) of the count of weekly bookings, denoted in equation 2.4.1.

$$\sinh^{-1}(y) = \ln(y + \sqrt{y^2 + 1}) \tag{2.4.1}$$

The IHS transformation has desirable properties similar to natural log, but its domain includes zero, which account for approximately 16% of our observations. For this draft we present the raw parameter estimates and loosely interpret them as percent difference. Like the non-linear transformation natural log, parameters alone do not have a clear interpretation. Norton (2022) provides details on the methods to transform parameter estimates into marginal effects.

2.4.3 Covariates

For the primary specifications, we include market-week fixed effects. The market-week fixed effects account for the fact that booking patterns could be different in different MSA's and that our data cover a period of rapid growth of the platform. Additionally, we include controls for number of reviews at the start of the week, appointment type (illness, new patient visit, or cross-listed), number of locations, appointment length, appointments offered, and hospital affiliation.

2.4.4 Sample Selection

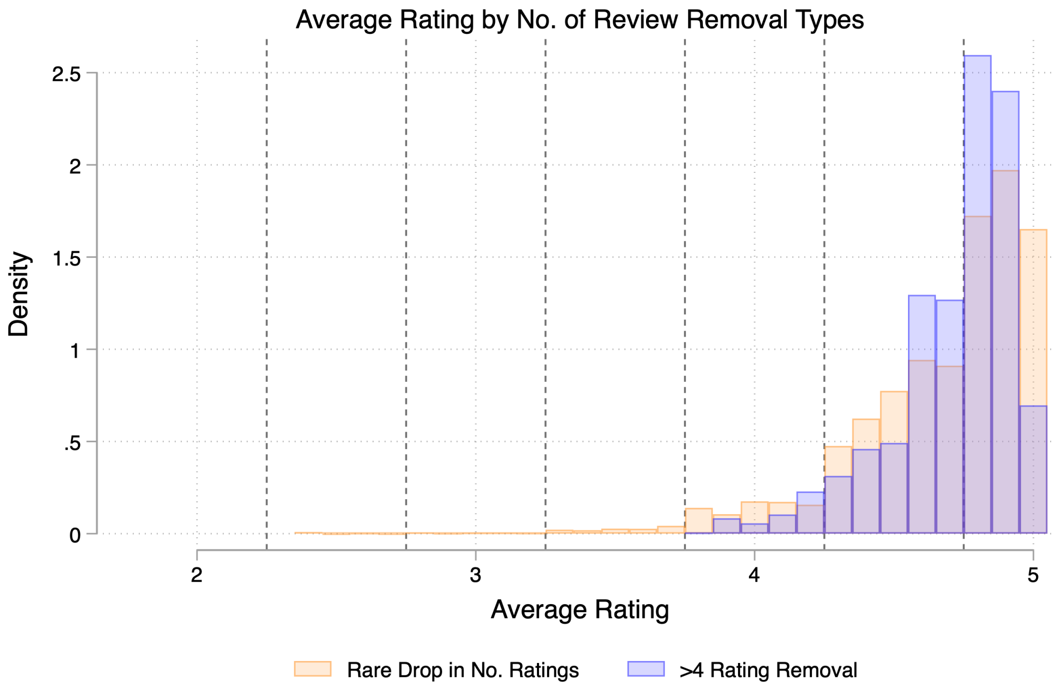
For our primary specification, the observations are on the physician week level. We limit our sample to appointments offered on weekdays between 8am and 6pm local time. We keep physician weeks where at least one appointment is available three weeks in advance. Our identification depends on cross-sectional variation in ratings near the cutoff values. Including observations with a large change in rating or were rating frequently change could be problematic for a number of reasons. For example, this type of within individual variation, a change in rating, depends on receiving new ratings, which requires new bookings, which is our outcome measure. Additionally, it is possible that the true underlying treatment effect of ratings accumulates over time. To account for this issue, we limit our observations to physician-weeks where the physician is at a "stable" half-star rating, meaning this is their half-star rating for at least 90% of included weeks.[5] We further limit our specification to exclude physician-weeks where the physician has fewer than eight ratings to avoid numerical issues in the running variable.

We make one other notable sample restrictions. We remove observations from physicians where there is evidence on possible deleted reviews or possible data collection issues related to number of ratings. If reviews are never deleted, then for a given physician the number of reviews should be monotonically increasing. We remove observations for physicians with

[5] Rating changes are rare, in our sample ∼$82% of observations are at a "stable" rating, and ∼$77% of observations are associated with physicians that who never change half-star rating during our period.

more than four weeks where the number of ratings decreased from the previous week. Figure 2.4.3 shows the distribution of physicians by average rating and review removal type. The figure below documents the distribution of observations flagged for deleted reviews.[6]

Figure 2.4.3: Appointments by Average Rating and Review Removal Type



2.4.5 Differential Impacts

In our analysis, the regression discontinuity estimate τ_j is a weighted average treatment effect of potentially heterogeneous treatment effects. In our case, this heterogeneity is not only expected but is of primary interest. We investigate the differential impact of star rating by partitioning our data into economically interesting subgroups and estimating treatment effects separately by cutoff and subgroup. Our current analysis focuses on physician gender, number of ratings, and hospital affiliation.

2.5 Results

In this section we first present data visualizations relating the appointment level likelihood of booking and patient volumes. Then we present estimates for patient volumes and remaining

[6]See **Appendix Deleted Reviews** for supplementary analysis on deleted reviews.

vacant appointments. We end with the differential impact of ratings. For this version we primarily focus on the 5-star cutoff.

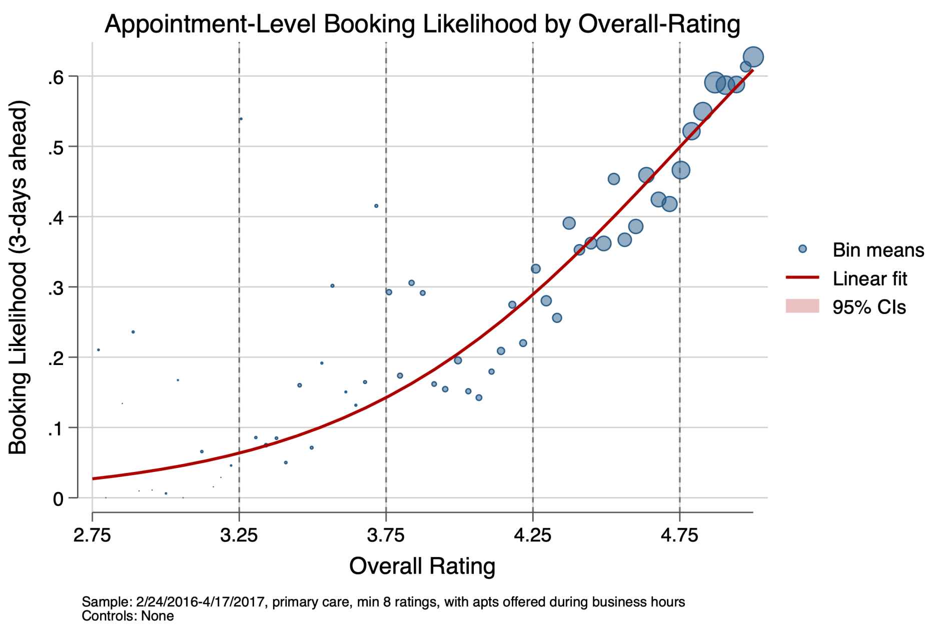
2.5.1 Naïve Approach Visualizations

These results show how likely a given appointment is to be booked by average rating, and then examine the CDFs of booked appointments over the number of days in advance that they are booked. Then we zoom into the 4–5-star physicians and present patient volumes by overall rating and star-rating.

2.5.1.1 Appointment Level Booking Likelihood

On the appointment level, as illustrated in Figure 2.5.1 we see a clear correlation between booking likelihood and average rating both across and within half-star cutoffs.

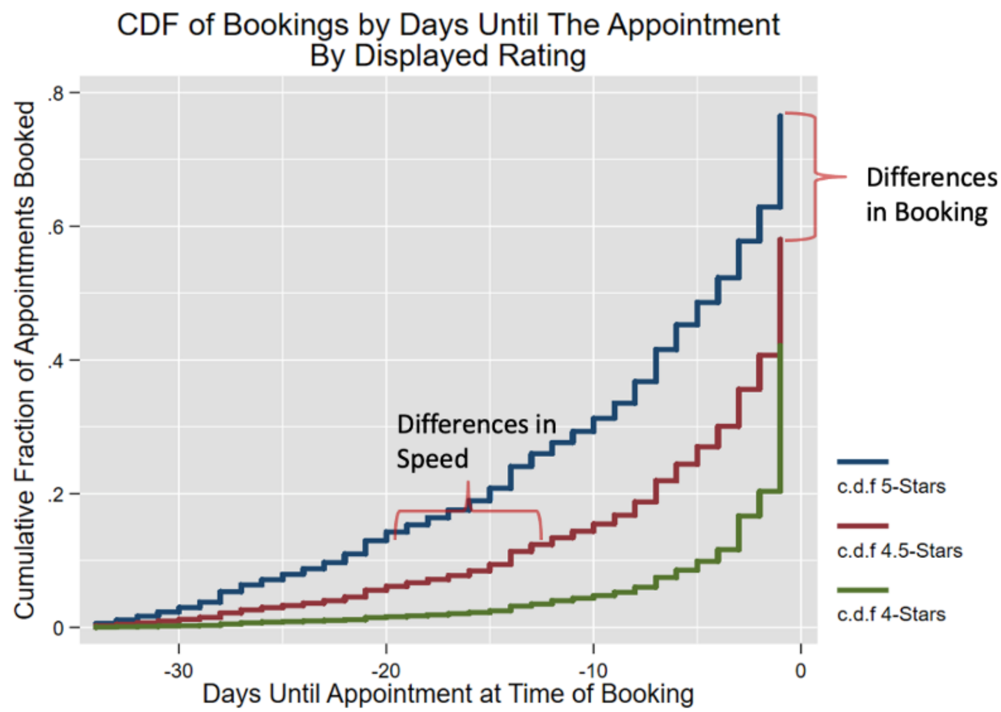
Figure 2.5.1: Booking Likelihood by Overall Rating



2.5.1.2 Cumulative Booking Likelihoods by Days Before the Appointment Date

Here we explore, by overall rating, when appointments are booked and how likely appointments are to be booked. Figure 2.5.2 presents a the fraction of appointments that are booked by time of booking. The horizontal axis is day of booking minus the day of the appointment, telling us the number of days before an appointment. The furthest in advance that an appointment could be offered is 35 days in advance. The CDF runs from 35 days before an appointment (left) to the day before an appointment (right).

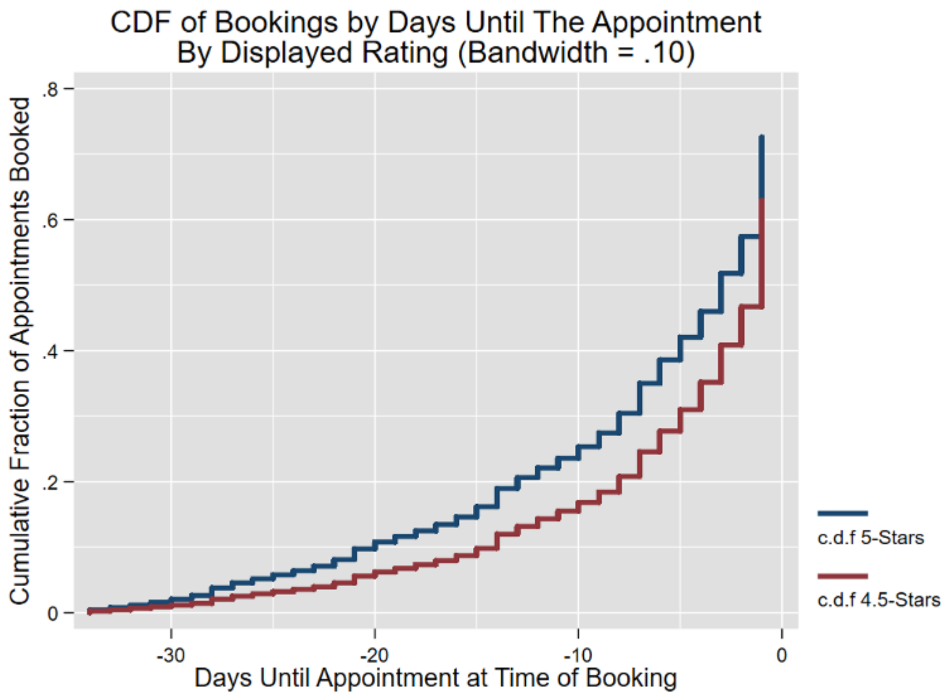
Figure 2.5.2: Comparison of Booking likelihood by Days in Advance of Appointment



Two useful ways to compare the CDFs are differences vertically and horizontally. The vertical difference tells us the difference in percent of appointments booked at a given number of days in advance. The horizontal differences can be interpreted as the difference in how many days in advance the same percent of appointments were booked. The CDF shows notable results. First, we see that higher rating doctors fill a higher percent of their appointments and do so further in advance. Second, we see that a high percent of appointments are booked the days immediately before the appointment date. About one third of 4.5-star bookings are done the day before, and over half the 4-star bookings happen the day prior. As mentioned above, this could be partially attributed to physicians removing vacancies from their schedules just before the appointment date.

Next, in Figure 2.5.3, we use a pilot bandwidth of .1 to compare these CDFs of physicians just above and just below the 4.75 threshold to have five stars. Here we see that the CDFs are closer together, but there are still differences in the percent of appointments booked and how far in advanced the appointments are booked. Here we see that the CDFs are closer together, but there are still differences in the percent of appointments booked and how far in advanced the appointments are booked.

Figure 2.5.3: Comparison of Booking likelihood by Days in Advance of Appointment near 5-Star Cutoff

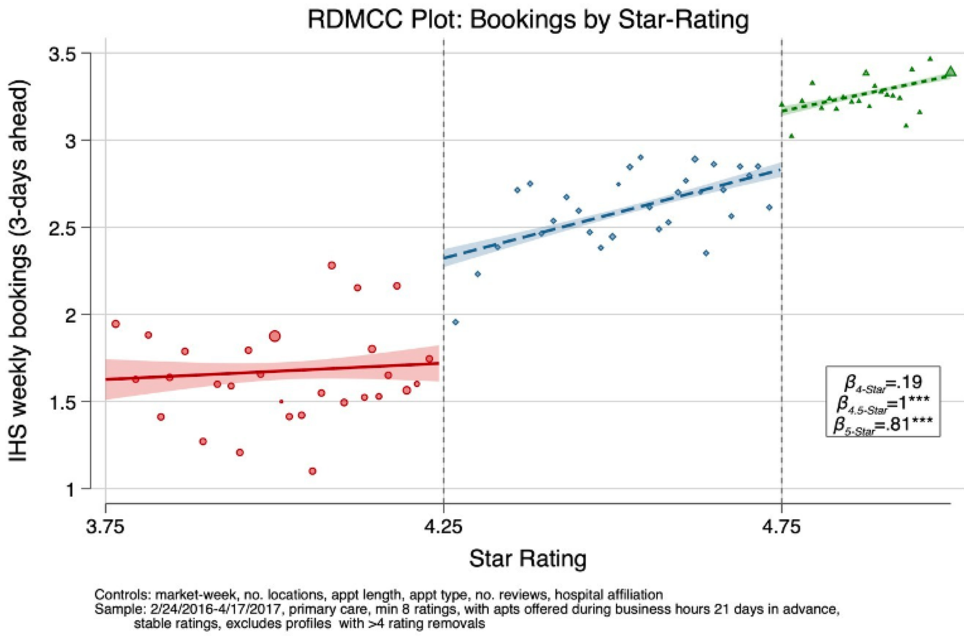


We next aim to quantify how much of these differences can causally be attributed to having 4.5 stars displayed versus having 5-stars displayed. However, one challenge highlighted by these figures is that the difference in booking likelihoods differs by time in advance of appointments. One possibility is that physicians have fixed capacity, so lower-rated physicians could receive more bookings closer to the time of appointments once appointments with 5-star doctors become scarce. One might underestimate consumer responsiveness to ratings if only looking at bookings just before an appointment. To address this concern, we repeat our analysis for different times leading up to appointments.

2.5.1.3 Booking Volume

While the previous figures focus on appointment level differences booking rates, we next consider booking volumes. Figure 2.5.4 plots weekly booking volumes by overall rating and displayed half-star. The horizontal axis is the average rating, and the vertical is the IHS of bookings aggregated to the week level. The jump at each cutoff presents suggests a jump in patient volume at the 4.5 and 5-star cutoffs. This figure is only suggestive, as it still takes a naïve, linear regressions approach instead of the data-driven regression discontinuity methods we present next.

Figure 2.5.4: Regression Discontinuity with Multiple Cumulative Cutoffs for Booking Volumes (4-5 Stars)



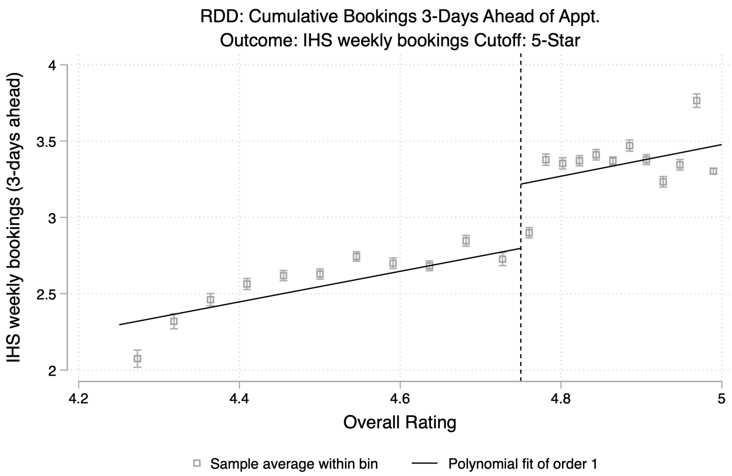
2.5.2 Regression Discontinuity Results: Impact of 5-Stars on Booking Volume

Figure 2.5.5 presents our RD results for the local average treatment effect of a five-star versus four-and-a-half star rating. Figure 2.5.5 presents the results for the weekly booking volume measured three days before the target appointment date. The left part of the figure displays the Regression Discontinuity plot, using the average overall rating as the running variable and booking volume as the variable on the vertical axis. On the right, the table shows the RD estimates calculated using conventional, bias-corrected, and robust methods.

We primarily rely on the robust result for our estimate. Our analysis identifies a substantial treatment effect of .761, which roughly approximates to a doubling of booking volumes. This finding underscores the influence of star ratings on patients' booking decisions.

Next, we assess the treatment effect across various booking windows. We repeat our analysis for different cutoffs of days ahead of the appointments, considering the cumulative booking volume leading up to these dates. As depicted in Figure 2.5.6, the leftmost estimate represents our regression discontinuity (RD) estimate for cumulative bookings made at least

Figure 2.5.5: RDD Results Cumulative Bookings Three-Days Aheah of Appt.



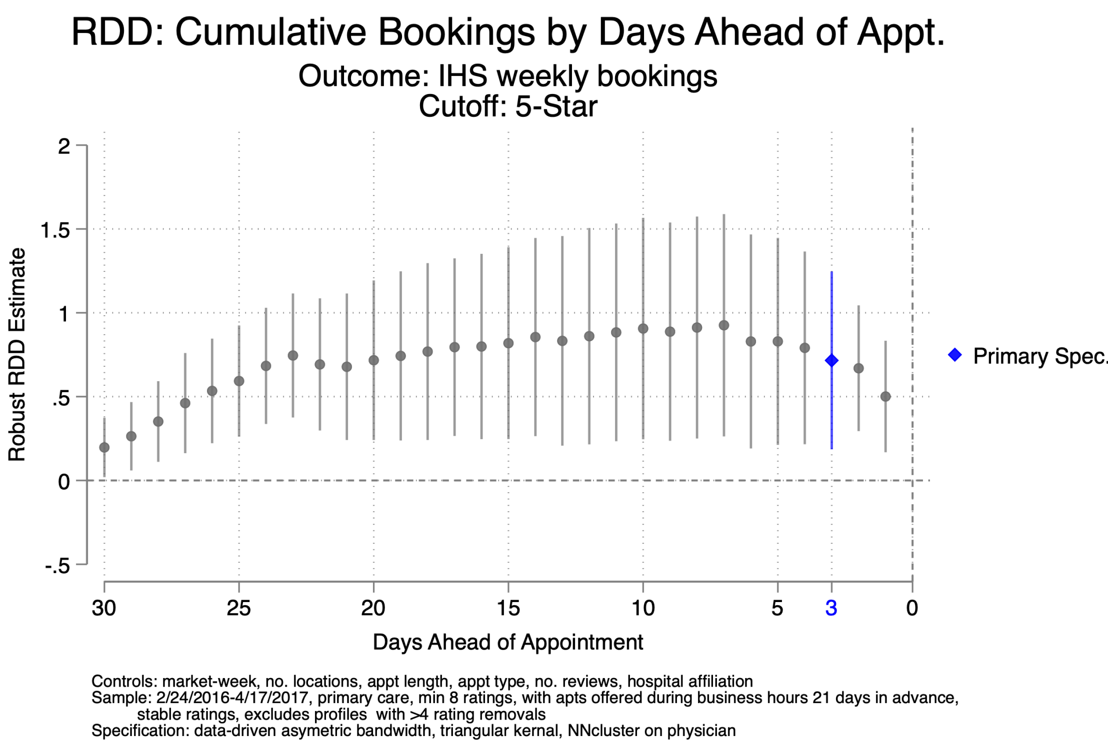
RD Method	**(1) IHS(Weekly Bookings)**
Conventional	0.686***
	(0.228)
Bias-corrected	0.716***
	(0.228)
Robust	0.716***
	(0.271)
Observations	54,263
Controls	✓

Controls: market-week, no. locations, appt length, appt type, no. reviews, hospital affiliation. **Sample:** 2/24/2016-4/17/2017, primary care, min 8 ratings, with appts offered during business hours 21 days in advance, stable ratings, excludes profiles with ¿4 rating removals. **Specification:** data-driven asymmetric bandwidth, triangular kernel, standard errors use NN clutters by physician

30 days ahead of the appointment dates and the rightmost estimate corresponds to cumulative bookings made up to one day before the appointment.

Our findings suggest that the treatment effect builds up until about a week before the appointment dates, then diminishes in the final week. One plausible explanation for this trend could be capacity constraints. Consumers show a clear preference for 5-star physicians, but as these top-rated physicians reach capacity and demand outstrips availability, 4.5-star physicians begin to receive bookings.

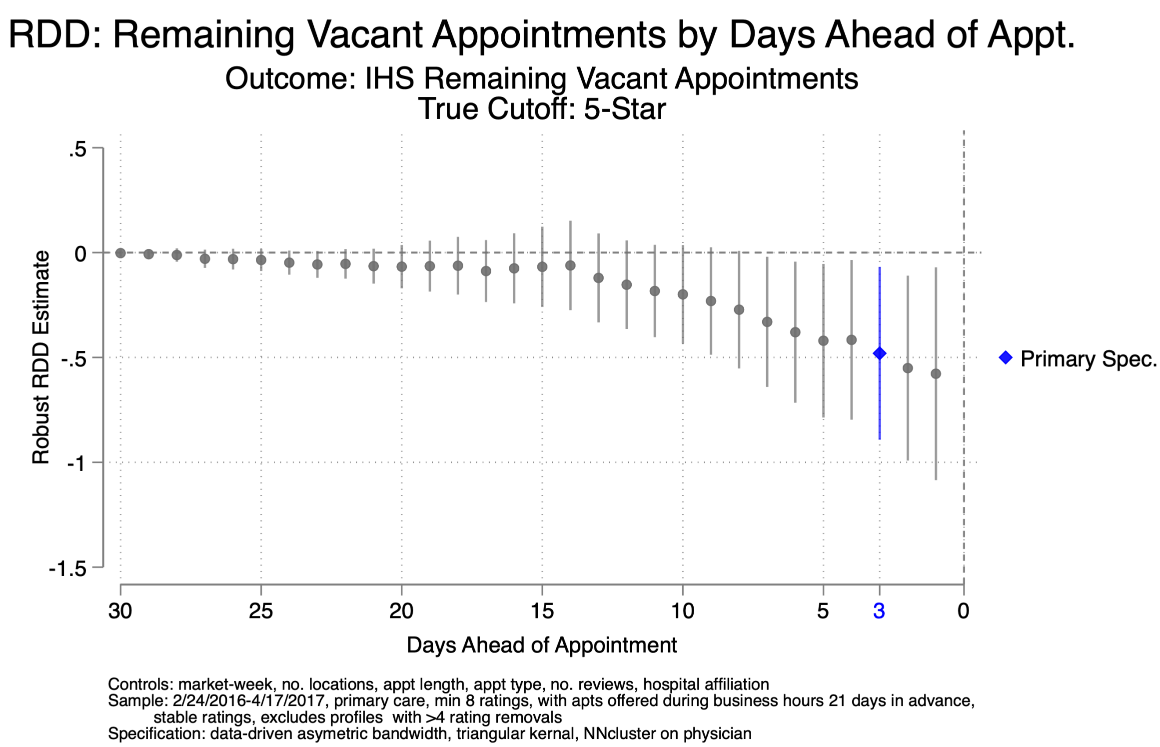
Figure 2.5.6: Booking Volume RD Estimates by Days in Advance of Appointment



2.5.2.1 Alternative Measure: Impact of 5-Stars on Vacant Appointment Volume

As an alternative measure of demand, we also analyze remaining vacant appointments. As physicians are capacity-constrained, measuring remaining capacity could also be a reasonable outcome. At the 4.75 cutoff, we find an effect where, compared to 4.5-star physicians, physicians with five stars have approximately 40% fewer vacant appointments three days before the appointment date. Figure 2.5.9 shows RD estimates from 30 days before the appointment dates to the day before the appointments.

Figure 2.5.7: Vacant Appointment Volume RD Estimates by Days in Advance of Appointment



2.5.3 Differential Impacts

Next, we evaluate the differential impact of ratings by partitioning our data into subgroups and repeating our analysis within each. In this section, we maintain our comparison between five-star and four-and-a-half-star ratings, focusing on weekly booking volumes measured three days ahead of the appointment dates. We concentrate on three comparisons of interest: physician gender, number of ratings, and hospital affiliation.

Our findings reveal that treatment effects are most pronounced for women physicians and those with a high number of ratings. There is a slight, but insignificant difference based on hospital affiliation, with a higher treatment effect noted for hospital affiliated physicians.

It is important to note that these results come with a caveat. While Regression Discontinuity Design estimates can have a causal interpretation under the right conditions, the differential impact might not necessarily hold this interpretation. Differences in parameter estimates could arise from heterogeneous treatment effects on other features, which might correlate with gender, number of ratings, or hospital affiliation.

2.5.3.1 Differential Impacts by Physician Gender

Figure 2.5.8 includes the RD plots of cumulative bookings three days ahead of appointments for women and men physicians separately, using the 5-star rating as the cutoff. The plots suggest an increase in booking volumes at the 5-star threshold for both genders. The figure also shows that women physicians have higher booking volumes relative to similarly rated men on both sides of the 5-star cutoff.

Figure 2.5.9 reports the RD estimates for booking volumes at the 5-star cutoff, separated by gender, and using three different estimation methods: conventional, bias-corrected, and robust. For reference, the figure also includes the pooled "base" estimates. The results indicate large significant effects on patient volumes for all subgroups and methods, though some results for men are significant at the 10% but not 5% level. Converting these point estimates to percentage changes, men see approximately a 106% higher booking volume, while women see a larger, approximately 276% higher booking volume.

Figure 2.5.8: Regression Discontinuity Plot of Booking Volume by Physician Gender

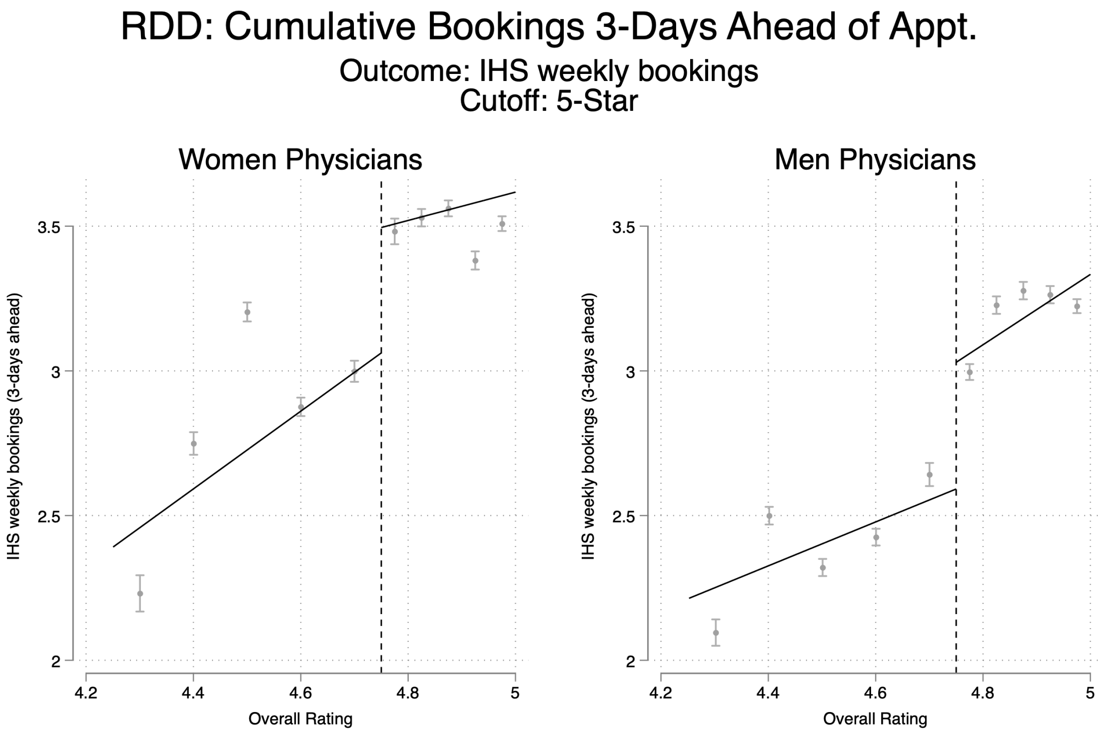
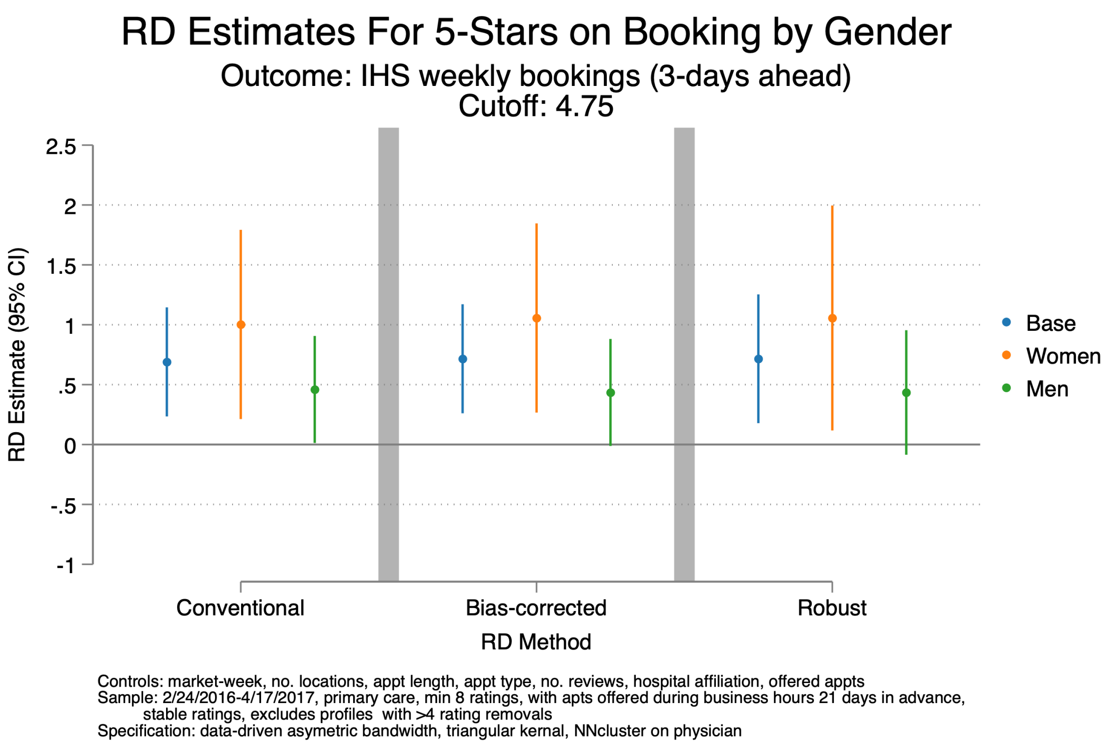


Figure 2.5.9: Regression Discontinuity Coefficients of Booking Volume by Physician Gender

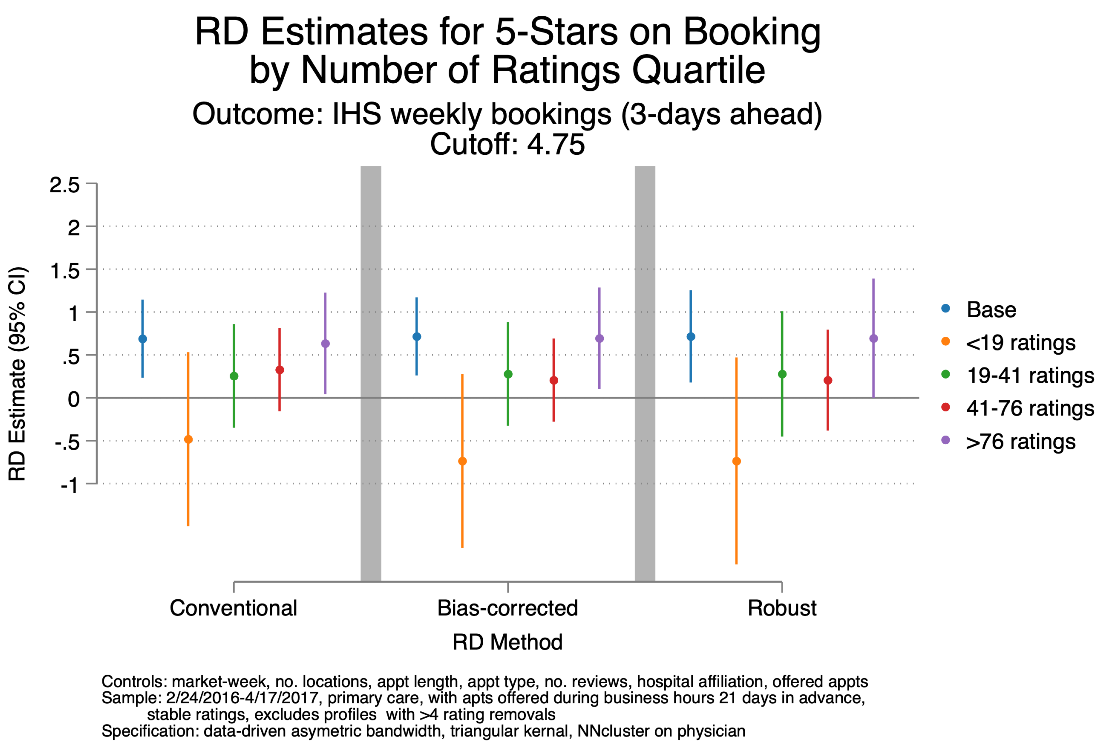


2.5.3.2 Differential Impacts by Number of Reviews

Next, we report the differential impact of rating by number of reviews. We included the number of ratings as a control in our primary specification to account for the fact that the number of ratings on its own could impact booking volume directly as a signal or indicate tenure on the platform or popularity. However, this does not rule out the possibility that the impact of ratings is different by number of reviews. For example, Luca (2011) finds different effects of ratings by number of ratings in restaurant markets. Luca (2011) suggests this could be due to Bayesian learning, where each review is a signal of quality, and more positive reviews, at the same rating, provide a stronger signal of quality. This results in a greater impact of average rating when there are more reviews.

Figure 2.5.10 reports the coefficient plot for cumulative bookings three days ahead of appointments, separated by quartile of number of ratings. The smallest quartile includes physicians with less than 19 ratings, and the largest includes physicians with more than 76 ratings. We find that the effect is greatest and most significant for physicians in the highest quartile (¿ 76 ratings). The results are not significant for the other quartiles. One of the possible drivers of this result is that consumers interpret ratings in a manner similar to Bayesian learning, whereby average rating has a greater impact when there are more ratings.

Figure 2.5.10: Regression Discontinuity Coefficients of Booking Volume by Number of Ratings

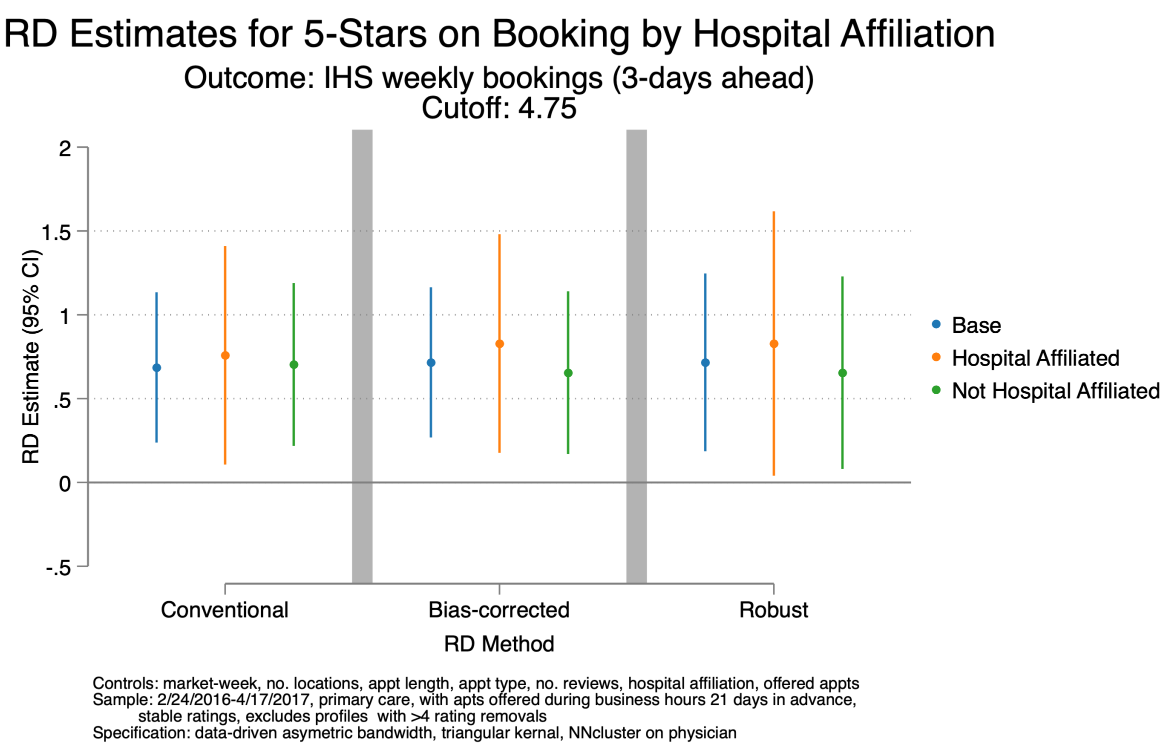


2.5.3.3 Differential Impacts by Hospital Affiliation

Next, we investigate the differential impact of ratings by hospital affiliation. Hospital affiliation could influence patient booking behavior as it may signal higher quality of care or access to better resources. We included hospital affiliation as a control in our primary specification to account for its potential direct impact on booking volumes. However, the impact of ratings may vary between hospital-affiliated and non-affiliated physicians. For example, patients might perceive ratings differently based on the perceived quality or reputation of the hospital affiliation.

Figure 2.5.11 reports the RD estimates for booking volumes at the 5-star cutoff, separated by hospital affiliation and using three different estimation methods: conventional, bias-corrected, and robust. For reference, the figure also includes the pooled "base" estimates. The results indicate significant effects on patient volumes for both hospital-affiliated and non-affiliated physicians across all methods. Although the magnitude of the effect appears to be slightly larger for hospital-affiliated physicians, there is no significant difference between hospital-affiliated and non-affiliated physicians. The results suggest that online reputation is a driver of booking volumes for both groups.

Figure 2.5.11: Regression Discontinuity Coefficients of Booking Volume by Hospital Affiliation



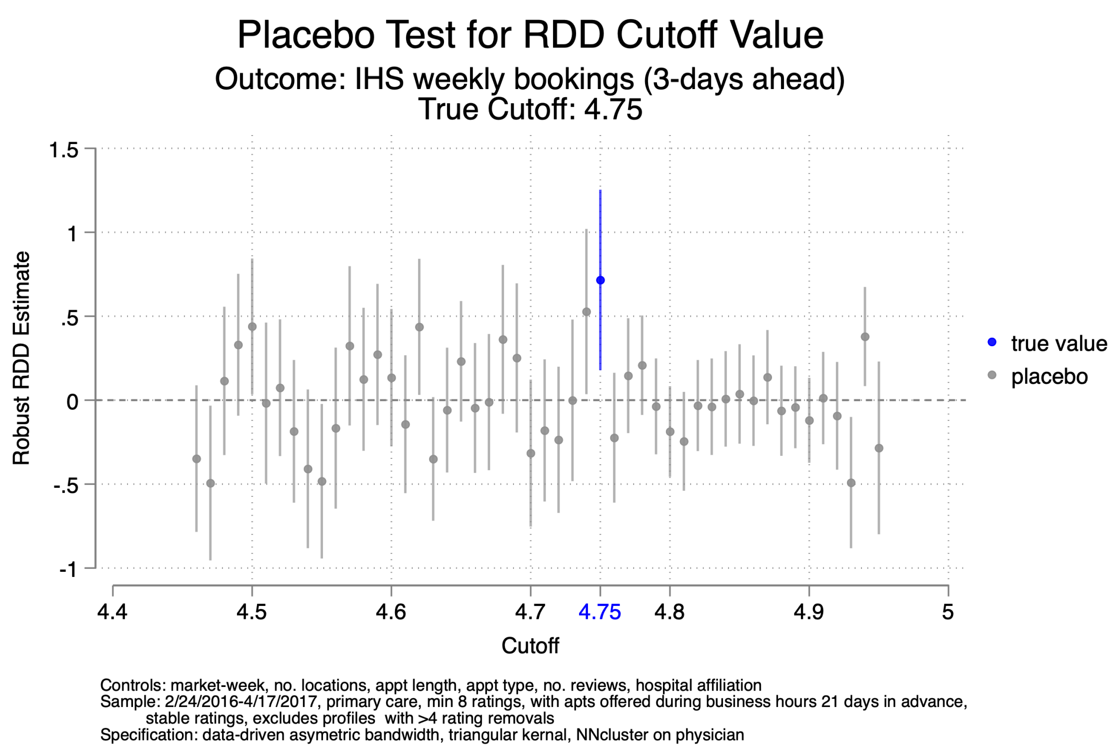
2.6 Robustness

In this section we investigate the robustness our primary specification using a placebo test and by investigating the distribution observations near the 4.75 cutoff.

2.6.1 Placebo Tests of Overall Rating at Alternative Cutoffs

As a placebo test, we repeat the RD analysis for specifying the discontinuity at alternative points other than the true cutoff. We find the greatest treatment effect estimate at the true cutoff of 4.75. There are however some points, for example 4.5, that would give false positive results.

Figure 2.6.1: Placebo Test of RDD Cutoff Value

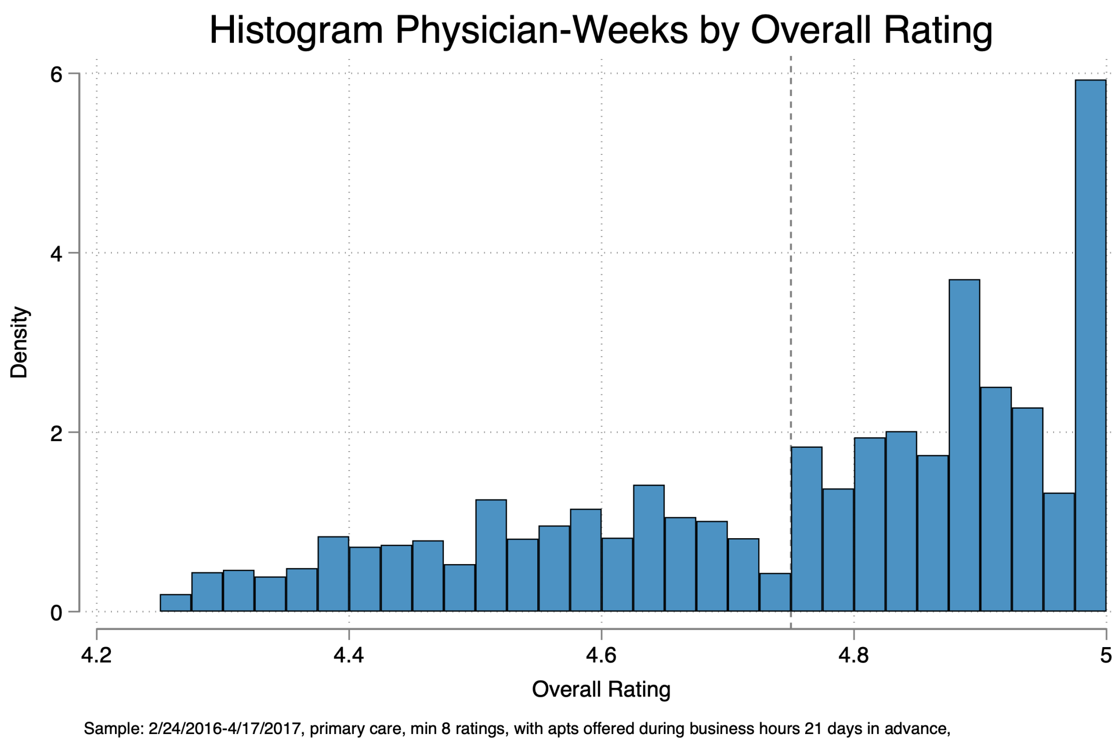


2.6.2 Rating Manipulation

A common concern when implementing RD methodology is gaming. In this setting, a physician that is near the overall rating discontinuity and knows that there is a discontinuous rating system may game the system to get their overall rating above the discontinuity. It is worth noting here that Zocdoc's closed loop rating limits the possibility of fake reviews. Only patients who have been verified to have visited the physician after booking an appointment through Zocdoc are encouraged to leave feedback. However, a physician may still encourage favored patients to submit positive reviews, or dispute negative reviews with the platform. As documented in the appendix, we do in fact find evidence of deleted reviews.

If gaming is driving the results then one would expect overall ratings to be clustered just above discontinuities. Figure 2.6.2 shows the density of physician-weeks by overall rating. We do see some clustering just above the discontinuity. The next step is to test the possibility that gaming is biasing the results. This would involve implementing the density test from McCrary (2008).

Figure 2.6.2: Placebo Test of RDD Cutoff Value



2.7 Discussion

We find positive and significant effects of ratings on patient volume, and different impacts depending on gender and number of ratings. A number of mechanisms could drive these results. For example, patients could have correlated preferences for physician gender and for ratings, and these preferences could also be correlated with frequency of going to the doctor.

Fink et al. (2020) find that patients have same gender preferences for both men and women physicians. Zocdoc reports that two in three of their users on the patient side are female.[7]

We find that ratings have the greatest impact for physicians with many rating. These results on the differential impact of overall rating by number of reviews are consistent with Bayesian Learning, were the influence of average rating increases with the number of signals, in this case ratings.

2.7.1 Platform Mechanics

It is important to also consider how Zocdoc's recommendation system might mediate the impact of ratings on patient volumes. In a search model, consumer choices would hinge on both their preferences and search costs, the latter depending on positions on the page. If the recommendation system ranks physicians based, at least in part, on average or rounded ratings, the effect of these ratings on patient volumes would depend not only on consumer preferences for ratings but also on how these ratings affect search costs through the recommendation system.

Moreover, depending on the recommendation system's training, the system's sensitivity to ratings could vary for different searches. This factor introduces another layer of complexity when assessing the relationship between ratings and patient volumes.

2.7.2 Next Steps

This subsection briefly discusses the next steps of this paper along four dimensions: data and analysis, robustness, and framing.

2.7.2.1 Additional Data and Analysis

The current analysis focuses on the impact of the 5-star rating threshold on booking volumes. A valuable next step would be to expand this investigation to other rating cutoffs, such as 4-star or 4.5-star thresholds. This analysis would provide a more comprehensive understanding of how different levels of ratings affect patient behavior. By examining multiple cutoffs, we can identify whether the estimated effects are unique to the 5-star threshold or if they generalize across various rating levels. This can also help to understand if there is a nonlinear relationship between ratings and booking volumes. However, as noted in Section 2.3, most physicians have high ratings. Unlike other online reputation platforms such as Yelp, where it is typical to see a normal distribution of average ratings, on Zocdoc, ratings tend to be very

[7] 12 Facts about Zocdoc Users

high. There is not much data below the 4.5-star cutoff, so one might worry about selection issues when evaluating differences in low ratings.

Beyond star ratings, the text of reviews can offer a deeper understanding of patient perceptions and the aspects of care that influence their decisions. Natural language processing (NLP) methods, including sentiment analysis and topic modeling, could reveal common themes and sentiments expressed in the reviews. The text of the reviews is noteworthy in three ways: 1) the direct impact of the text of reviews on booking volumes, 2) how the text of reviews changes the impact of ratings, and 3) what the text of the reviews can tell us about the information contained in ratings. Summarizing the differences in the sentiment of reviews by rating could inform us how much online ratings convey information about actual quality.

This paper evaluates the differential impact of ratings by gender, number of ratings, and hospital affiliation. As next steps, we plan to include additional comparisons by economically interesting groups and groups whose results would inform us about the underlying mechanisms driving the impact of ratings on booking volume. For instance, examining the differential impacts of ratings by apparent demographics inferred from profile photos could reveal biases or preferences related to demographics. Comparing subgroups based on specialty or education might uncover mechanisms driving the observed effects. Additionally, repeating our analysis for different booking windows, as we do in section X, could inform us about some of the mechanisms driving the large difference in booking and treatment effects by subgroups.

2.7.2.2 Dynamic Incentives and Physician Behavior

We find that the ratings, and by extension the rating system, influence which physicians consumers choose. Our results, some of which show more than doubling booking volumes, suggest a considerable incentive to receive good ratings. While potentially outside this project's scope, these findings open questions about how physicians might adjust behavior in response to the rating system. For example, do physicians take actions to improve ratings? If so, are these welfare-improving actions, such as improving quality of care? Do they include actions like deleting or contesting bad ratings?

2.7.2.3 Additional Robustness Checks

The paper would benefit from expanding the robustness checks and including those standard in the regression discontinuity design literature. A sensitivity analysis would repeat our analysis with different sampling rules and RD settings. A balance test would check for differences in physician characteristics across the cutoff. Additionally, a formal test for bunching around the rating thresholds, such as the McCrary (2008) test, might indicate strategic behavior by physicians to improve ratings to get just above a respective cutoff. Given the apparent mass point just above the five-star cutoff, one approach is donut regression discontinuity, where we would drop all observations near the cutoff. These robustness checks will help ensure the reliability and robustness of the conclusions drawn from the analysis.

2.7.2.4 Framing

There is also room to improve this paper to better frame it at the intersection of digital markets generally and healthcare specifically. In many settings, consumers use online ratings to choose experience goods such as hotels, restaurants, and movies, or retail settings with ex-ante, but not ex-post, uncertainty about product quality. Physician services are a credence good with ex-ante and ex-post uncertainty about product quality. On one hand, one might expect ratings to be less important in consumer decision-making since online ratings might be less informative about quality, as those leaving reviews are not fully informed about the quality of care they receive. On the other hand, consumers might appear even more sensitive to ratings since they have even less ex-ante information about medical services. Our planned analysis that looks at the impact of ratings in the presence of other quality information, as well as analyzing the text of reviews, could improve the paper in this direction.

2.8 Conclusion

Using data collected from Zocdoc, this paper analyzes a unique data set containing physicians' appointment schedules, professional information, and reviews from verified patients. Zocdoc displays ratings on a scale of one to five stars, with overall average ratings rounded to nearest half star. Since ratings are rounded to the nearest half star, we use a regression discontinuity framework to identify the causal impact of patient reviews on patients' choice of physician. On the appointment level, our results indicate that a half star improvement in displayed rating means that appointments are more likely to be booked and also booked further in advance. On the physician-week level we find that 5-star physicians have higher patient volume, and that these results are most pronounce for women physicians, and physicians with many ratings. We test the robustness of our model using alternative specifications and placebo tests.

CHAPTER 3

Long-Term Echoes of Short-Term Policy: Tracing the Persistent Impact of Medicare Advantage Subsidies

with Thomas Buchmueller, William Mandelkorn, and Sarah Miller

3.1 Introduction

Governments intervene in healthcare markets through a variety of regulations and subsidies, to encourage health insurer participation and align payer incentives with social or program goals. Such policy interventions directly affect insurer and consumer incentives when the policies are in place. However, these policies may also generate effects that persist well after the policy phases out. This persistence may be especially present in health insurance markets, where previous researchers have shown that consumers exhibit high degrees of inertia.[1] If present, such long-running effects are important to document because they alter the possible benefits–and costs–of such interventions.

We provide new evidence on the long-lasting effects of government subsidies on health insurance markets by analyzing the interaction of two policies in the Medicare Advantage (MA) market. Medicare Advantage (MA) is a $203 billion program in which the federal government contracts with commercial insurers to provide coverage for Medicare beneficiaries. Over time, the levels and nature of these per enrollee payments, called "benchmark" payments, have changed. In this paper, we examine the impacts of these higher benchmark payments on MA market outcomes and the prices for medical care, along how these impacts change as these higher payments are largely phased out.

[1]For example, Strombom et al. (2002), Handel (2013), and Abaluck and Gruber (2016). For a wider ranging review, see Handel and Kolstad (2015).

To examine this question empirically, we use two sources of policy variation. First, we analyze a discontinuity in capitated Medicare Advantage payment rates first exploited by Duggan et al. (2016). In that paper, the authors document that MA payment rules increase discontinuously for counties associated with a metropolitan statistical area (MSA) with population exceeding a 250,000 threshold. This results in otherwise similar counties that fall close to this threshold being exposed to very different MA payment rates, potentially influencing the number of plans introduced in these markets, the fraction of beneficiaries who enroll in an MA plan, and the bargaining power of the plans themselves with respect to hospitals and providers in the area. Second, we expand on this variation to examine how a later policy change in the Affordable Care Act (ACA), which greatly reduced this gap at the 250,000 MSA population threshold, affected MA enrollment and plan participation. We also analyze a novel MA county price index, which we previously developed in Buchmueller et al. (2022) and derived from data on provider transaction prices provided by the Health Care Cost Institute (HCCI). We use this price index to investigate whether higher benchmark payments affected the prices paid by MA plans for outpatient and inpatient care.

Consistent with previous work, we find that counties that received higher benchmark payments had significantly higher enrollment in Medicare Advantage plans compared to otherwise similar counties that just missed the population cutoff. And, in a novel contribution, we show that these higher benchmark payments also affected market prices: MA plans in counties that fall just above the population cutoff pay lower prices for outpatient care compared to plans in counties just below the cutoff, although inpatient prices appear largely similar for MA plans in these two sets of counties. We hypothesize that these price differences reflect the fact that MA plans above the population cutoff covered more patients and therefore had greater bargaining power vis a vis outpatient providers compared to plans below the cutoff.

Then, we document that between 2012 and 2016, the discontinuity in benchmark payments was largely phased out as the result of ACA policy changes. By 2016, county benchmarks on either side of the population cutoff were indistinguishable, and the difference in payments after accounting for plan quality bonus payments and the MA bidding process had fallen by more than half. However, we find that markets exposed to higher benchmark payments in the past remained significantly different across several dimensions even after higher payments were reduced. These counties continued to have higher rates of MA enrollment, greater MA penetration and more participating MA plans long after the payment advantage was greatly reduced. Indeed, even 5 years after these higher payments began to be phased out, the discontinuity in these outcomes at the population cutoff was essentially unchanged compared to the period when much higher payments were in place. Similarly, we continue to find significantly different outpatient prices in counties that were initially exposed to the higher benchmark payments, even after this payment differential is reduced. These price effects are particularly large in markets where the insurance side is highly concentrated. These results demonstrate that large subsidies can shape long-run insurance market outcomes even after these subsidies are reduced. Given the large reduction in the difference in payments across the cutoff from 2012 to 2016 and the continued elevated enrollment, these results may point to the relative importance of the number of covered lives in insurer-provider price and network negotiations.

3.2 Background

3.2.1 Medicare Advantage

Medicare beneficiaries have a choice of two types of public health insurance. Beneficiaries can enroll in traditional Medicare (TM), in which the federal government operates as an insurer and sets parameters such as cost-sharing and reimbursement rates. Alternatively, beneficiaries can select from a menu of private health insurance plans, called Medicare Advantage (MA) plans, which may have different premiums, cost-sharing arrangements, benefits, and provider networks. Private companies offer Medicare Advantage plans but are paid a capitated payment from Medicare for each beneficiary they enroll.

Historically, this per enrollee payment to MA plans was based on observed medical costs in traditional Medicare in that plan's county, as measured using previous years' spending in traditional Medicare. However, to avoid having counties in which this payment was considered inappropriately low, Medicare introduced a uniform floor in 1998. This floor resulted in higher payments for counties with historically lower spending in Medicare. Starting in 2001, Medicare established separate floors for "urban" and "rural" counties, with urban counties receiving a higher floor payment. Medicare considered a county to be urban if it was part of a metropolitan statistical area (MSA) with 250,000 residents or more. As a result, counties associated with MSAs just above and just below the 250,000 population threshold experienced very different benchmark floors despite being otherwise similar in terms of their underlying input costs and population needs. In 2004, the Medicare Modernization Act and newly defined metropolitan and micropolitan statistical areas brought about additional changes in benchmarks. Benchmarks were raised, and a county's urban classification and floor status were recalculated based on these new MSA delineations. Additionally, counties that qualified as an urban county in 2001 were grandfathered in to urban status, regardless of their 2004 MSA populations.

The Medicare Modernization Act additionally introduced a competitive bidding system for insurer payments. In 2006, this competitive bidding process was implemented, in which, after county benchmarks are published, insurers make a bid for each plan and segment (which may span multiple counties), based on expected enrollment and expected risk scores amongst expected enrollees in that plan-segment. This bid represents the insurer's estimate of the cost to cover Medicare Part A and B benefits for an enrollee, and may include administrative costs and profit capped at 15% of the total plan revenue.[2] These bids are adjusted by CMS based on expected risk and an expected benchmark for each bid is calculated. When MA eligibles enroll in a MA plan, the plan-segment bid is adjusted by the ratio of the enrollee's county benchmark to the expected plan-segment benchmark. If the insurer's bid is above the benchmark, insurers are required to charge enrollees a premium to cover the difference. If the insurer's bid is below the benchmark, insurers receive a portion of the savings as a rebate. Prior to 2012, insurers received 75% of the savings as a rebate with which they were required to provide extra benefits. After 2012, this rebate percentage transitioned to 50 - 70% of the savings, dependant on the contract's quality star rating. Research indicates that benchmark payments and realized payments track each other closely, as plan bids increase on average by about $0.50 for every $1 increase in the county's benchmark (Song et al., 2013).

In 2010, the Affordable Care Act made a variety of changes to MA reimbursement intending to bring MA benchmarks more in line with costs in traditional "fee for service" Medicare. In the "post-ACA" period, benchmark calculations would transition to a new formula, consisting of the product of county per capita TM spending and both county and plan percentage adjustments. These ACA changes resulted in an end to the urban and rural floor policies. First, benchmarks for 2011 were frozen at their 2010 level. After 2011, yearly benchmarks were subject to a cap equal to the counterfactual pre-ACA benchmark for that county-year. Counties transitioned from the pre-ACA to post-ACA benchmark calculation over a set period of 2, 4, or 6 years, the length of which was determined by the difference between the county's Projected 2010 Benchmark (a one time calculation based on the county's 2010 fee for service (FFS) rate) and the county's 2010 pre-ACA rate,[3] with counties receiving a longer transition period for larger differences between the two rates. As there is no difference between the post-ACA benchmark calculation formulas for urban and rural counties, we should expect the difference in benchmarks for urban and rural counties to close entirely by the end of the transition period in 2017.[4] This policy change additionally provides the opportunity to observe the "long run" impact of historically high rates (among urban counties with MSA

[2]https://bettermedicarealliance.org/wp-content/uploads/2020/03/BMA_WhitePaper_MA_Bidding_and_Payment_2018_09

[3]https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Downloads/Advance2012.pdf

[4]While the post-ACA benchmark calculation policy does not factor in MSA population thresholds or floor amounts, the prior floor policy could have some residual impact through the use of pre-ACA benchmark caps.

populations above the 250,000 cutoff) while the difference between these rates narrows over the post-ACA period.

Additionally in 2012, the ACA instituted plan quality bonus payments, aimed at incentivizing increases to plan quality. Each plan could receive an additional payment on top of the county benchmark amount based on that plan's quality star rating. Plans with a rating exceeding certain star thresholds could receive a 3-5% bonus payment, with this bonus doubling in certain counties with low FFS spending and high MA enrollment. While these bonuses were uncapped for the first few years of the policy change, the sum of benchmarks and quality bonus payments were capped at the pre-ACA benchmark rate beginning in 2015.[5]

A variety of studies have taken advantage of pre-ACA variation in MA payments to explore how payment increases affect consumers and plans. For example, Cabral et al. (2018) and Duggan et al. (2016) both investigate whether increased payments affect enrollment in MA, and to what extent payment increases are passed on to consumers in the form of lower premiums and enhanced benefits. Both studies document incomplete pass through, i.e., when payments increase, less than the full amount is passed on to consumers in the form of lower premiums or enhanced benefits. These results suggest that plans in MA markets have substantial market power. Similarly, Song et al. (2013) find that a $1 increase in Medicare benchmark payments increases plans' bids by $0.53, consistent with the result that MA plans exercise market power. Other studies have used variation at the 250,000 MSA population cutoff to investigate the relationship between MA enrollment with hospital admissions and mortality (Afendulis et al., 2017) and opioid abuse (Baker et al., 2020; Rhodes, 2020).

While these studies present relevant evidence on how MA policy affects plans and patients, less is known about how the prices received by providers might respond to changes in MA payments and enrollment. Descriptive evidence shows that, for inpatient care, average MA and TM prices are not very different. For example, Curto et al. (2019) present statistics comparing hospital payments for patients in MA and TM and find that, within a hospital and diagnosis related group, payments from MA plans are only about 1 percent higher than what is received by the hospital for seeing a TM patient. However, prices for ED care are 9 to 10 percent higher in MA as compared to TM. The authors speculate that this similarity in prices across TM and MA for hospital services may reflect the regulation that requires hospitals to accept the TM prices for beneficiaries if the hospital is not included in the MA network. This regulation may allow MA plans to bargain down hospital prices close to the rate paid by traditional Medicare. Other research has documented similar patters; for example, Maeda and L (2018) also show similar prices paid by MA and TM and further demonstrates a negative

[5]For more information on the exact bonus payment percentages by year, please see yearly CMS Advance Notices.

correlation between MA enrollment and the ratio of MA to TM prices. Baker et al. (2016) similarly use HCCI inpatient claims data to show MA plans pay 5-8% less than FFS Medicare for hospital admissions. They additionally document a negative relationship between prices and MA penetration rates, showing that MA plans in high penetration rate CBSAs pay lower prices than MA plans operating in CBSAs with lower penetration rates.

Additional studies have illustrated several other potential avenues through which MA prices may vary. Xu and Polsky (2023) analyze MA and TM prices using Hospital Price Transparency Rule data, finding MA and TM outpatient prices deviate substantially at roughly half of the hospitals studied. These deviations often result in higher prices for MA plans, and the authors find that higher markups are more often found at hospitals in rural areas. Lin et al. (2022) compare MA and FFS Medicare prices for dialysis using claims data from HCCI, finding that higher payments are made by MA plans and higher markups are charged to MA plans by larger dialysis chains. The authors additionally suggest that these large dialysis chains may leverage their market power during negotiations with MA plans to bargain for these higher prices. Meyers et al. (2018) examine differences in quality of care for MA and FFS Medicare enrollees, specifically with regards to skilled nursing facilities (SNFs). They find that FFS Medicare enrollees are more likely to enter higher-rated SNFs, and attribute some of the difference in likelihood of high-quality SNF admission to the fact that MA plans are more likely to have narrower provider networks in an attempt to reduce MA plan costs.

These studies suggest that inpatient hospital prices are already low in MA as compared to other commercial insurance, but are higher for hospital outpatient care. Increasing the payments plans receive for providing beneficiaries coverage, and the resulting increase in plan enrollment, could shift these observed prices. Additional levers that may affect bargained for prices, like differences in provider location, high amounts of market power, or likelihood of network inclusion, may affect not only MA prices with respect to FFS Medicare prices, but may affect prices across MA plans as well. In the next section, we describe one potential mechanism through which payments could affect transaction prices: altering bargaining between plans and providers.

3.2.2 Why Might a Subsidy Continue to Have Effects While Being Phased-Out?

There are a variety of reasons why a subsidy that is greatly reduced could continue to generate long-lasting changes in insurance markets. For example, a large and growing literature demonstrates that there are large switching costs and consumers display substantial inertia in health insurance markets (Handel, 2013; Marzilli Ericson, 2014; Nosal, 2011). If a subsidy spurs higher enrollment in MA, then high inertia may result in continuing effects even after the subsidy phases out. Furthermore, some research suggests that entry of health insurance plans into a market may be driven in part by the fixed cost of entering (Geddes, 2022). If, for example, substantial fixed costs arise from the initial setup of provider networks, plans may incur these fixed costs to enter while the subsidy is in effect, and it may not be profitable to exit even when the subsidy that induced that entry has begun to erode.

3.2.3 How Could Subsidies Affect Prices?

In addition to studying how these higher benchmark payments affect enrollment in MA plans over time, we also present novel evidence on the role of these subsidies in market prices. We hypothesize that benchmark payments may affect prices in the short- and long-term through several channels. First, Gowrisankaran et al. (2015) (hereafter, GNT) and Gaynor et al. (2015) (GHT) frame the insurer-provider bargaining in a Nash Bargaining framework with a "take it or leave it" setup for prices and network inclusion. This Nash Bargaining setup is a useful theoretical framework to derive a more comprehensive understanding of how Medicare Advantage subsidies might influence the bargained prices for medical care. This section first summarizes the Nash-Bargaining setup. We then describe three mechanisms which we refer to as the incomplete-passthrough effect, the enrollment effect, and the entry effect, all of which may affect bargained-for prices.

3.2.3.1 Nash-Bargaining

The below equation illustrates the GNT model bargaining between managed care organizations (MCOs) and providers, using a Nash-Bargaining model with take it or leave it prices, meaning MCOs and providers bargain over provider system, insurer, procedure-specific prices to maximize the weighted product of the provider and insurer objective functions:

$$NB^{m,s}\left(p_{mj_{j\in J_s}}|\vec{p}_{m,\text{-}s}\right) = \underbrace{\left(\sum_{j\in J_s} q_{mj}(N_m, \vec{p}_m, e_m)[p_{mj} - mc_{mj}]\right)^{b_{s(m)}}}_{\text{Provider Surplus}} \underbrace{\left(V_m(N_m, \vec{p}_m) - V_m(N_m \setminus J_s, \vec{p}_m)\right)^{b_{m(s)}}}_{\text{Insurer Surplus}} \quad (3.2.1)$$

The Nash Bargaining solution maximizes equation 3.2.1 for each target provider-system, insurer, procedure specific price, holding all other prices and networks fixed.

$$p^*_{mj} = \max_{p_{mj}} NB^{m,s}\left(p_{mj}, p^*_{m,-s}|p_{m,-j}\right) \quad (3.2.2)$$

The key elements of this objective function, relevant to our discussion, include provider surplus and insurer surplus. Provider surplus depends on expected quantities and markup. Expected quantities, q_{mj}, depend on the insurer m's network, prices, and the number of enrollees. The markup is the difference between the target price and marginal cost. Insurer surplus, on the other hand, is the difference between $V_m(N_m, \vec{p}_m)$, the value to the insurer of including the target provider in the network with price vector $\vec{p}_m$, and $V_m(N_m \setminus J_s, \vec{p}_m)$, the value to the insurer of without the target provider in the network, keeping the remaining prices and network constant. The parameters b_s and b_m represent the bargaining weights, which sum to one.

To make the relationship from subsidy to enrollment to expected quantity more explicit, we slightly modified the notation from GNT and GHT by including number of enrollees, e_m, in the expected quantity function. Additionally, subsidies, enrollment, and market conditions may modify the relationship between V_m and the network and price choice variables. Although we do not explicitly model bargaining, if one were to do so, it could be beneficial to delineate the connection between enrollment and the network, which is currently embedded in the value function on the insurer surplus side.

How do these elements influence prices? Assuming we have an equilibrium with an interior solution, holding all else constant, policies that increase provider surplus would theoretically apply downward pressure on the target price. Conversely, policies enhancing insurer surplus would apply upward pressure on prices. With this setup, we can consider how varying subsidy levels influence the Nash Bargaining equilibrium. Specifically, we will consider the mechanisms through which subsidy would impact provider surplus, insurer surplus, or both.

3.2.3.2 Incomplete-Passthrough Effect

Consider a simple example of "incomplete-passthrough" of a subsidy. When the government partially subsidizes the cost of a good, firms may choose to capture a portion of the subsidy, by decreasing their price by less than the subsidy amount. This phenomenon is referred to as incomplete passthrough, since the full benefit of the subsidy is not passed on to the consumer.

In the context of Medicare Advantage subsidies, Duggan et al. (2016) find evidence of incomplete passthrough, with private insurers accruing much of the benefit of additional subsidies and a smaller share going to patients. We propose that this incomplete passthrough could also impact negotiated procedure prices. When negotiating prices, providers are aware of any subsidies being paid to insurers, and may attempt to bargain for higher procedure prices in an attempt to capture some of this surplus.

Applying this to the Nash Bargaining framework, all else equal, the subsidy raises the insurer's value $V_m(N, P, B))$ of including a target provider in their network which would lead to a higher Nash Bargaining price equilibrium.

3.2.3.3 Enrollment Effect

The next mechanism to consider is the enrollment effect, which relates to how variations in enrollment influence the bargaining dynamics. Changes in subsidy levels could affect the number of MA enrollees. An increase in enrollees would enhance an insurer's bargaining position, as providers could gain from a larger patient pool. Within the Nash-Bargaining framework, more enrollees leads to increased expected profits for providers through higher expected patient volumes. This dynamic would theoretically exert downward pressure on the equilibrium prices, given that providers anticipate higher overall revenue from the expanded patient pool.

3.2.3.4 Entry Effect

We describe the entry effect as the mechanisms through which subsidies encourage new insurers to enter the market, and the corresponding influence of additional insurers on negotiated prices. Given a partial passthrough of the subsidy through lower prices or higher plan quality, insurers could attract more enrollees. However, new insurers may be incentivized to enter the market since additional subsidies partially cover insurer costs.

On the provider side, with more insurers in the market, providers may have an opportunity to play insurers off each other to secure better terms in bargaining. In the Nash-Bargaining framework, with enrollees divided among a larger number of insurers, expected quantities per insurer could be lower, which, in equilibrium, would put downward pressure on prices.

On the insurer side, it is worth noting that, in the Nash-Bargaining setup, we could see changes in insurer surplus. For example, more insurers could put competitive pressure on plan prices, plan quality, and network size. The direction of the entry effect on negotiated prices is unclear since some of these forces could be offsetting. All else equal, one would expect lower plan prices to decrease negotiation prices, similar to a reversal of the incomplete-passthrough effect. Conversely, if overall demand is more sensitive to network size, one would expect upward pressure on negotiated prices.

3.2.3.5 Combined Mechanisms

The incomplete passthrough, enrollment, and entry effects, while not exhaustive, underscore the potential ways subsidy could influence bargaining dynamics. In markets with incomplete passthrough, providers may extract some of this windfall for themselves through negotiations in the form of higher prices. However, subsidies could impact enrollment and entry. In the

Nash-Bargaining setup, each of these forces would influence equilibrium negotiated prices. To better understand these mechanisms, we separately estimate the impact of benchmark payments on enrollment, insurer entry, market concentration, and negotiated procedure prices. To the extent that price differences are driven by enrollment effects, we would expect these differences to persist even after the subsidy begins to erode if enrollment remains elevated in plans in the affected counties.

3.3 Data

3.3.1 MSA Definitions and Populations

We take advantage of the urban floor cutoff used in Duggan et al. (2016) and others combined with variation from later changes in this policy due to the Affordable Care Act to investigate the effect of changes to government payments made to insurers on insurance markets. Crucial to this approach is a discontinuity in MA benchmarks for urban counties that occurred at an MSA population of 250,000. We take advantage of this discontinuity in a regression discontinuity design approach, with MSA population serving as the running variable that determines each county's rural or urban status. Due to policy changes that occurred in 2001 and 2004, we will need each county's MSA population for both 2001 and 2004.

From the US Census Bureau, we begin with county-year population estimates for both 2001 and 2004,[6] along with city and town population estimates for 2001.[7][8] We additionally use two historic MSA delineation files: 1) the MSA delineation file for Dec 2003; and 2) the MSA delineation file for 1999. These files define MSAs for 2004 and 2001 respectively. For each county, we then calculate their associated MSA population for both 2001 and 2004. Both populations are needed, as counties that qualified for urban status in 2001, but not 2004, were grandfathered in as urban counties.[9] We use these two measures to create our running variable, what we refer to as the "relevant MSA population." This measure is the MSA population that determined a county's urban status, and is equal to the county's 2004 MSA population for non-grandfathered counties and the 2001 MSA population for grandfathered counties.[10]

[6] https://www2.census.gov/programs-surveys/popest/datasets/2000-2009/counties/totals/

[7] https://www2.census.gov/programs-surveys/popest/datasets/2000-2009/cities/totals/

[8] While MSAs are delineated by counties in 2004, in 2001, some New England counties were broken into multiple MSAs at the town-level.

[9] These grandfathered counties generally arise due to two possible scenarios: 1) the MSA population has dropped between 2001 and 2004; or 2) the county was assigned to a different MSA with a lower population in 2004.

[10] As we use population estimates that change over time, there may be some measurement error inherent in this variable. We believe only one county may have received the wrong urban status assignment due to this

3.3.2 MA Benchmark, Enrollment, Market Penetration, and Plan Data

MA benchmarks are the maximum monthly capitated payments made to MA plans per enrollee. While these benchmarks were originally paid in full to insurers, the amount paid to insurers in later years was adjusted by both plan quality bonuses and the competitive bidding process. We examine multiple aspects of this payment process. We first derive a measure of the "base" MA benchmark from yearly CMS rate calculation data books.[11]. For each year, we use the risk-adjusted benchmark rate before any budget neutrality adjustments.

In later years, this base benchmark was adjusted in two major ways. As part of the ACA policy changes, benchmarks were adjusted by each plan's star rating, with higher quality plans receiving bonus payments of up to 10% on top of the county benchmark. Using Medicare Part C and D Performance Data,[12] we first derive the star rating of each contract. We then calculate the plan quality bonus benchmark, equal to the sum of the county benchmark and contract-county quality bonus payment, for each contract-county using the contract star rating, the aforementioned CMS rate calculation data, and yearly CMS Announcement and Advance Notice documents.[13] We then use the county-contract enrollment data, described below in more depth, to calculate contract enrollment-weighted average county star ratings and bonus benchmarks.

Realized payments to insurers were further adjusted through the competitive bidding process, first implemented in 2006. While these bids somewhat track county benchmarks, we use MA bid pricing data[14], rate calculation and bonus benchmark data, Medicare Part C and D performance data, and county-plan-level enrollment data to derive enrollment-weighted average county bids, savings, rebates, and premiums. By examining payments to insurers through multiple channels, we are able to examine three ways government payments to insurers may change: 1) via changes to county benchmarks; 2) via potential additional quality bonus payments due to higher plan quality; and 3) via potential changes to payments resulting from changes to insurer bids.

Our analysis focuses on a subset of comparable counties whose 2004 FFS rate is below the rural floor (referred to in Duggan et al. (2016) as "group 1" counties). Within these counties, all counties to the left of the urban cutoff receive the rural floor for their 2004 MA benchmark, while all counties to the right receive the urban floor. By limiting to these counties, we can

error, and we drop that county from our analysis.

[11] https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Ratebooks-and-Supporting-Data

[12] https://www.cms.gov/medicare/health-drug-plans/part-c-d-performance-data

[13] https://www.cms.gov/medicare/payment/medicare-advantage-rates-statistics/announcements-and-documents

[14] https://www.cms.gov/medicare/payment/medicare-advantage-rates-statistics/data

examine the full effect of the higher benchmark payments given to urban MA plans, as all counties in the affected group should have similar FFS costs, but only a subset receives higher payments. We use the 2004 FFS rate found in the 2004 MA rate calculation data book to identify these counties.

We collect MA enrollment data from two CMS sources. For 2008 through 2021, we use MA county penetration data,[15] while we use State County files for 1997 through 2005.[16] As the later data is provided monthly, and the earlier data is provided quarterly, we use estimates for MA enrollment and MA eligibles from March for both sources.[17] For later files, MA enrollment is censored at 11 enrollees per county. For our analysis, we censor the earlier data using an 11 enrollee cutoff as well. We use data on MA enrollees and MA eligibles to calculate the MA penetration rate, a measure of market penetration equal to the number of enrollees divided by the number of eligibles in each county.

Additionally, we use contract-level enrollment and service area data from CMS to construct measures of insurer competition at the county-year-level. We begin with monthly enrollment data at the county-plan-contract-level[18] and impose two restrictions on the sets of available contracts. First, we use monthly contract service area data[19] to restrict the set of counties within which a contract may operate.[20] Second, within each month, we limit to county-plan-contracts that have greater than 10 enrollees. We then construct three county-year-level variables to measure insurer concentration and market participation. We first calculate the number of plan-contracts within a county-month that have greater than 10 enrollees. We then calculate the number of parent organizations that offer these plans within each county-month. Lastly, for each county-month, we compute the number of enrollees for each parent organization, enrollment-based market shares for each parent organization, and county-month HHIs. We then average these variables within each year to arrive at our county-year estimates.

3.3.3 Medicare Advantage Price Indices

To construct our outcome variable, we rely on a price index generated from MA claims data recorded in the Health Care Cost Institute (HCCI) database. A detailed description of these price indices, and direct access to the indices themselves, is found in our concurrent paper Buchmueller et al. (2022). Here, we briefly describe the methodology we used to produce

[15]https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MCRAdvPartDEnrolData/
[16]https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/HealthPlanRepFileData/SC
[17]Due to data availability, we use June estimates for 1997 and 2008.
[18]https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MCRAdvPartDEnrolData/
[19]https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MCRAdvPartDEnrolData/
[20]We do this to limit any "movers" in the enrollment data, those who may enroll in a contract, but later move to a new location not covered in their contract's service area.

these indices, and the following discussion borrows heavily from that paper.

The HCCI data contains detailed information on inpatient, outpatient, and pharmaceutical claims for over 50 million commercially insured individuals per year. Previous research has used these data to study health care spending and prices in a variety of settings including prescription drugs, dialysis, pediatric hospitals, and acute care hospitals (see Cooper et al. (2018), McCarthy and Raval (2022), Dafny et al. (2022), League et al. (2022)). In addition to recording data on the general commercially insured population, the data contains significant coverage of the MA population, with coverage of about 40 percent of all MA enrollees.[21]

To construct the price indices, we generate two datasets comprised of procedures and discharges to patients whose payer is Medicare Advantage observed between 2012 and 2016. The first dataset is at the outpatient procedure level and contains over 235 million procedures over the five years that the data cover. The second dataset includes information reported at the inpatient discharge level. These data include roughly 7 million inpatient discharges. Within each of these discharges, we observe the patients' diagnoses and the procedures performed during the hospital stay. Within each dataset, we estimate county price indices by regressing the log of the price of the procedure or discharge on county by year fixed effects and a set of control variables.[22] For the outpatient regression, we include as controls procedure by year fixed effects, the log of the number of "units" of the procedure received by the patient (e.g. a claim item of 60 minutes of physical therapy may be billed as four 15 minute units), and patient demographics (e.g., age group and gender). For the inpatient regression, we include diagnosis related group (DRG) by year fixed effects, the log of the length of stay, and patient demographics. We also construct alternative versions of each index with other combinations of controls, as well as separate versions of the index that rely on only procedures or hospital stays that occur in versus out of network. These alternative constructions allow us to assess the sensitivity of our analyses to the exact empirical specification we use to generate the indices. From these sets of regressions, we obtain the coefficients on the county fixed effects to use as price indices.

In order to conduct our analysis, we make two sample restrictions. First, we restrict our sample of counties to those whose fee for service rates are such that they would be bound by the rural floor if their associated MSA population fell below 250,000.[23] For these counties, we expect their position relative to the 250,000 MSA population threshold to have a meaningful effect on benchmark payments. Second, we are only permitted to disclose estimates from the HCCI database if the number of observations support exceeded 10. Our analysis is therefore

[21]See https://www.healthdatamanagement.com/articles/hcci-gets-full-access-to-medicare-claims-data.

[22]We identify county by mapping the zip code of the provider to the county.

[23]These are referred to as "Group 1" counties by Duggan et al. (2016).

limited to counties and years that meet this disclosure requirement.

3.4 Empirical Approach

To analyze the impact of higher benchmark payments on enrollment, plan variety, and prices over time, we take advantage of the interaction between two policies. First, we use a regression discontinuity design that exploits the discontinuity in MA benchmark payments for counties associated with MSAs with 250,000 population or higher. Second, we take advantage of the fact that the ACA changed this discontinuity and reduced the payment difference at the cutoff, while leaving the enrollment differentials largely unchanged. The transition between these policies allows us to examine how the market impacts change over time, as the higher payments to insurers are gradually phased out and the discontinuity in benchmark payments is reduced.

We operationalize this approach by estimating a regression discontinuity design model at the 250,000 population cutoff by year. We use the same MSA population bandwidth (from 150,000 to 350,000 population) used in Duggan et al. (2016), with a triangular kernel. We further weight our data to place higher weight on county level observations associated with greater precision of the price index by weighting by the inverse of the variance of price index estimate. To conduct inference, we estimate using cluster-robust nearest neighbor variances (Abadie and Imbens, 2008). In our preferred specification, we include covariates measured prior to the implementation of the benchmark floor, the log of the 1997 fee for service rate and the log of 1997 Medicare Advantage enrollment, to improve precision. However, our analysis is very similar if we exclude these control variables. We estimate these models annually to capture the policy-driven changes in the discontinuity over time and because the price indices are comparable within, but not across years.

3.5 Results

3.5.1 How did Payments to Insurers Evolve Over Time At the Population Cutoff?

We first examine how the floor policies affect monthly benchmark payments that MA plans receive over time. First, Figure 3.6.1 plots county benchmark payments relative to the 250,000 population threshold for five years broadly representative of the five "eras" of the floor policy. First, Panel (A) plots average benchmark payments in 2000, prior to the urban floor being instituted, against the county's relevant MSA population. In this "untreated" year, we find no evidence that benchmark payments differed across the 250,000 population cutoff. Panel (B) shows a similar figure using data from 2003, when the urban floor was in place. In this year, we see a small but significant change in benchmark payments at the cutoff. Panel (C) shows data from 2008, after the Medicare Modernization act greatly increased the size of the benchmark discontinuity at the cutoff. Here, we see that counties that fell just above the population threshold received substantially higher payments compared to those below. Panel (D) shows data from 2014, during the ACA transition that phased out the floor policy. In this year, the discontinuity in payments is much less pronounced than in Panel (C). Finally, Panel (E) shows benchmark payment averages from 2019, after the ACA transition was completed and the benchmark discontinuity was nearly eliminated. In this year, there is no longer an apparent discontinuity in benchmark payments at the population cutoff.

We combine these annual RD analyses into a single figure in Figure 3.6.2. In this figure, we show our estimated discontinuity in benchmark payments at the 250,000 population threshold by year. Vertical lines indicate the years of major policy changes and show the years in which we observe price data in HCCI. We see that, starting in 2001 and increasing in 2004, counties whose MSA population placed them right above the urban floor cutoff experienced significantly higher benchmark payments per month per enrollee than counties whose payments placed them right below that cutoff. The difference during this period was between \$60 and \$80 per month. Then, starting in 2012, the new ACA rules began to phase in which steadily closed the gap between counties on either side of the urban floor cutoff. By 2017, the difference in benchmark payments across the threshold was close to zero and not statistically significant. It remains statistically insignificant through 2021, the last year for which we have this data. This pattern is similar without controls; see Appendix Figure C.1.1.

While county benchmarks provide a base maximum capitated payment to insurers for enrollees within each county, realized payments to insurers are additionally affected by the quality of their offered plans. Dependent on plan quality and certain county characteristics, quality bonus payments implemented in 2012 offer potential increases to insurer payments. If an offered plan's star rating exceeds certain thresholds, plans are eligible to receive a quality bonus payment of up to 5% of the county benchmark. In certain counties, these quality bonus payments may be doubled to a maximum of 10% of the county benchmark. Even if there is no difference in county benchmarks across the urban cutoff, if insurers offer higher quality plans or enrollees enroll in higher quality plans in urban counties, insurer payments may be higher in urban counties due to higher quality bonus payments.

We study this potential change to insurer payments by examining how both plan quality and quality bonus payments change across the urban cutoff. In each year, after determining the star rating for each MA contract, we construct a contract enrollment-weighted average star rating within each county. If there are increases to this average star rating across the urban cutoff, this could indicate that either insurers offer higher quality plans in urban counties or that enrollees are more likely to enroll in higher quality plans in urban counties, both of which could result in additional payments to insurers across the urban cutoff. We examine this possibility in Figure 3.6.3. Similarly to county benchmarks, for each year, we plot the estimated discontinuity and 95% confidence interval for the county average contract star rating in panel (A), limited to Group 1 counties within the DSV [150k, 350k] bandwidth. With the exception of 2013, there is no significant difference in the enrollment-weighted average contract star rating across the urban cutoff.

We can similarly examine the difference in "bonus" benchmarks, the sum of each county's benchmark and the contract enrollment-weighted average quality bonus payment within each county, across the urban cutoff. In contrast to our results in Figure 3.6.2 which show the discontinuity in benchmarks reducing to an insignificant amount by 2016, annual estimates of the discontinuity in bonus benchmarks, shown in panel (B) of Figure 3.6.3, follow a decreasing pattern from 2012 through 2014, but eventually settle near a $25 discontinuity across the urban cutoff. Through 2014, while county benchmarks were subject to a pre-ACA benchmark cap, the quality bonus payments were not, and could allow the bonus benchmark to be in excess of the pre-ACA benchmark. This policy was changed, and for all years after 2014, the sum of the county benchmark and quality bonus payment was subject to the pre-ACA rate cap. Bonus benchmarks in rural counties exceed the pre-ACA rate cap more often than in urban counties, leading to a continued discontinuity in bonus benchmarks after 2015, despite no difference in the average contract star rating across the urban cutoff. While this result implies a continued possibility of higher payments to insurers after the 2012 ACA policy changes, the estimated discontinuity still reduces from near $70 to around $25, amidst a time period of large increases to county benchmarks overall.

Realized payments to insurers will also be affected through the competitive bidding process, in place since 2006, described in more depth above. We examine the effect of the urban cutoff on multiple aspects of the bidding process, namely insurer bids, CMS capitated payments to the insurer, and insurer rebates. We first use plan-segment bid and plan enrollment data to create yearly enrollment-weighted county average bids. We then compute a measure of CMS capitated payments to insurers, equal to the lesser of the benchmark and bid, and create yearly enrollment-weighted county averages of CMS payments. We additionally compute the rebate for each plan-segment-county, equal to 0 if the bid is above the benchmark and a portion of the savings if the bid is below the benchmark, and create yearly enrollment-weighted county average rebate amounts. We additionally compute the sum of CMS payments and rebates, and find yearly enrollment-weighted county average CMS payment and rebate amounts, meant to reflect the total amount given back to plans per enrollee, to fund either expected costs or additional benefits. We examine the effect of the urban cutoff on each of these variables.

In panel (A) of Figure 3.6.4, we show the yearly estimated discontinuity in average MA bids across the urban cutoff. Given prior research that shows MA bids closely track MA benchmarks (Song et al., 2013), it is unsurprising to see estimated discontinuities in MA bids that show similar patterns to estimated discontinuities in MA benchmarks. There is a broadly positive and increasing effect across the cutoff until 2012, after which this difference decreases to an insignificant amount by 2017. A similar pattern is observed in the yearly discontinuity estimates for average CMS payments as well, illustrated in panel (B) of Figure 3.6.4.

Panel (C) of Figure 3.6.4 shows the yearly estimated discontinuity across the urban cutoff for average rebate payments within each county. While there is a slight increase in estimated rebate payments across the urban cutoff prior to 2012, this effect becomes either insignificant or fairly small towards the end of our price index data sample. When examining the sum of CMS payments and rebates, as shown in panel (D) of Figure 3.6.4, we see patterns similar to those observed in our yearly benchmark and bonus benchmark estimates. We see increases across the urban cutoff that grow until the ACA policy changes take effect, after which these differences across the cutoff shrink to an estimate near $20.

MA county benchmarks have been subject to many policy changes over the years. In this paper, we focus mainly on the effects of a policy that gave higher payments to insurers in a subset of urban counties. While this policy began to be phased out with the implementation of ACA policies in 2012, remnants of past policies could be felt both as counties transitioned to new post-ACA policies and as certain payments remained subject to pre-ACA payment caps. While we observe that the urban cutoff has an insignificant effect on county benchmarks by 2016, realized payments to insurers are contingent on both additional potential quality bonus payments and the competitive bidding process. When examining the effect of the urban cutoff on county average bonus benchmarks or county average CMS payments and rebates after bidding is complete, while the effect of the urban cutoff diminishes after 2012, there is still a small and significant positive effect which persists after 2012. While this provides an avenue for insurers in urban counties to receive higher payments from CMS than those insurers in rural counties, this difference is decreasing across our main 2012 through 2016 sample period, and continues to decrease over time with respect to increasing county benchmarks.

3.5.2 Impact of the Cutoff on Enrollment, Plan Entry, and HHI

The urban floor and subsequent ACA payment policies had a substantial impact on the amount of payments a plan received per enrollee. We next use a similar empirical approach to examine how the MA penetration rate and the log of total MA enrollment evolved over this period. Our results are presented in Figure 3.6.6, with the MA penetration rate reported in panel A and log of MA enrollment reported in panel B.[24] Associated plots showing the averages of these variables by MSA population for the five years mentioned above are reported in Appendix Figure C.1.2 and C.1.3. We see significant differences in both MA penetration and log MA enrollment beginning in 2008. On average, counties falling just above the 250,000 population cutoff have an approximately 10 percentage point higher MA penetration rate in 2008, rising to about a 15 percentage point higher MA penetration rate by 2012. This estimate is slightly greater than the estimates reported in Duggan et al. (2016), perhaps indicative of a stronger effect in later years or cumulative effects on enrollment over time. However, in contrast to the difference in benchmark payments, we do not see any evidence that this difference in MA penetration rates becomes smaller when the ACA policy transitions out the higher benchmark payments for counties just above the urban floor. Indeed, our estimate in 2021 is 16.6, even larger than the 14.7 percentage point discontinuity we observe in 2012 when the discontinuity in benchmark payments was largest. In short, the impact of large benchmark payments on MA penetration outlived the higher payments themselves.

Higher benchmark payments may encourage plans to increase enrollment via advertising or lowering premiums. However, the fact that payment pass through is incomplete also implies that plan profit per enrollee has increased, which could encourage entry by other plans. We investigate whether the number of plans active in a county increases at the 250,000 threshold and report these results in Figure 3.6.7, which plots RD estimates by year beginning in 2007. We find an increase in the total number of plans offered of about 70% at the threshold. The effect remains significant through 2022, a full decade after the subsidies began to be phased out, although this effect does appear to become somewhat smaller over time. Since the same parent company can offer multiple plans, we also investigate the number of parent companies participating in a county changes at the threshold.[25] We again see a significantly higher number of participating parent companies just to the right of the population threshold, with counties above the threshold experiencing about 35% more participating parent companies than those just below the threshold. This effect on parent company participation remains significant after the ACA transition. Finally, we examine HHI, as measured using enrollment

[24] Recall we do not have data for 2006 or 2007.

[25] The same insurer can offer multiple plans; we refer to the insurer as the parent company (e.g., Aetna, Humana). We derive this variable from the plan-level enrollment data.

by parent company. As additional plans enter, competition in these markets should rise. However, we do not find a statistically significant difference in HHI at the population cutoff. This is primarily because enrollment in the marginally offered plan is low.

Overall, our results show that a large increase in the capitated MA subsidy to plans had long-term effects on enrollment in MA, MA enrollment as a percentage of total Medicare enrollment, and the number of plans operating in the county. These differences persist even 6 years after the full transition to a new benchmark policy that resulted in a complete removal of the higher benchmark payments and a sizeable reduction in realized payments to plans. Indeed, the effects of the subsidy on MA penetration and log enrollment hardly change in the years following, despite the subsidy diminishing greatly.

3.5.3 Impact of Payments on Outpatient and Inpatient Prices

We next examine whether transaction prices changed at the 250,000 population cutoff, and whether these changes demonstrated the same persistence over time reflected in the enrollment and plan participation data. We start by examining how in-network prices changed for outpatient and inpatient procedures. Figure 3.6.8 plots our county price index for in-network outpatient procedures against county MSA population. Marker sizes correspond to the weights given to these observations in our RD, with size weighted by the product of a triangular kernel and the inverse variance of the price index estimate. Each panel corresponds to a different year. We see relatively high values of the price index immediately to the left of the cutoff, with lower values to the right of the cutoff. This change at the cutoff is consistent across years. A similar analysis is conducted for inpatient visits in Figure 3.6.9. While the price index values appear somewhat lower for inpatient visits across the cutoff, the difference at the cutoff is visually small, except in 2016.

Figure 3.6.10 presents the associated RD estimates for in-network outpatient (panel A) and inpatient (panel B) prices in each year, with corresponding values reported in Tables 3.6.1 and 3.6.2. We see a statistically significant and negative impact of falling above the cutoff on outpatient prices of about 0.28 log points in 2012. This estimated effect remains stable across years, mirroring the stability in the effect of the cutoff on the log of MA enrollment documented in Figure 3.6.6. We observe a negative effect on outpatient prices, that remains present and stable even after the discontinuity in benchmarks across the cutoff began to greatly decrease. We ascribe these lower outpatient prices mainly to the enrollment effects observed at the cutoff, with MA penetration rates in urban counties maintaining a much higher level than penetration rates in rural counties throughout our sample. That is, MA plans operating in counties just above the cutoff population were able to extract lower outpatient prices due to the fact that they covered a larger number of individuals. This hypothesis is further bolstered by the fact that insurance markets do not appear to be less concentrated just above the cutoff despite serving a larger number of enrollees (Figure 3.6.7).

In contrast, the effect on the price of inpatient stays is quite small and not statistically significant, with the exception of 2016 in which we observe a significant negative impact on inpatient prices.

We additionally examine how these pricing effects change with respect to MA insurer concentration. One hypothesized mechanism through which payments may decrease prices is by improving an insurer's bargaining position by increasing the number of patients covered by that insurer. But, the marginal effect of additional enrollees may be less important in a market that is highly competitive. To explore this relationship, for each year, we first categorize counties into one of four insurer HHI quartiles, based on the prior year's MA parent organization county-level MA enrollment shares. We then re-run our yearly price index RD within each HHI quartile and plot the results for in-network outpatient procedures and inpatient stays in Figure 3.6.11. While the patterns may differ subtly across years or categories, our results generally illustrate that within higher insurer HHI county groupings, price effects become more negative as the MA market within counties becomes more concentrated. While price effects are generally insignificant for the counties with the lowest insurer HHIs, price effects are more strongly negative for those counties within the highest insurer HHI quartile. These results could help emphasize the large impact bargaining may have on prices for MA care. When the MA insurer market is less concentrated, insurers may struggle to leverage higher payments or higher enrollments into lower bargained-for prices with providers, who may play these many insurers off of each other. However, as the MA insurer market becomes more concentrated, these insurers may be more able to leverage higher enrollments across the urban cutoff for larger price decreases with providers.

3.5.3.1 Robustness of Price Index Estimates

We discuss the development of our MA price index estimates in Buchmueller et al. (2022). To examine the robustness of our price index results to other specifications, we run a series of robustness checks for other price index specifications and RD bandwidth choices. Additionally, given the difference in slopes that can be observed in Figure 3.6.8, we run checks to ensure our results are not being driven by a small number of observations to either side of the 250,000 MSA population cutoff, but are instead likely resulting from this urban cutoff.

Our price index RD results are robust to a number of alternative ways of specifying the RD model and constructing the price index, which we demonstrate in Figure 3.6.12. For each year, we conduct our estimates using alternative versions of the price index model. For outpatient procedures, we constructed versions of the price index that account for patient demographics and/or modification codes in the claims files. For inpatient stays, we have constructed alternative price indices that control for log of the length of stay and/or patient demographics. For each of these different versions of the price index, we report a version of the RD estimates using both a data-driven bandwidth (denoted with a dashed line) and the Duggan et al. (2016) bandwidth (solid line). The top panel shows the results for the outpatient price index and the lower panel shows the results for the inpatient price index. Across all versions, the results are quite stable.

We also conduct our analysis on subgroups of procedures and inpatient stays, reported in Figure 3.6.13. For outpatient procedures, we examine ER visits, ambulance services, radiology services, and lab pathology services. For inpatient care, we analyze stays associated with mental health, surgical care, and skilled nursing facilities (SNFs). Within these categories of procedures and hospital stays, we find largely similar patterns as when we examine price effects overall.

We additionally conduct a series of donut RDs to examine how those observations nearest to the 250,000 MSA population cutoff may drive the results seen in Figure 3.6.10. In Figure 3.6.14, we estimate donut RD effects for both outpatient and inpatient in-network prices. We estimate these effects for a variety of donut sizes, for both those Group 1 counties within the Duggan et al. (2016) bandwidth and all Group 1 counties using a data-driven bandwidth. Across specifications, for small donut sizes, we see our estimated effects are relatively consistent, indicating that the effects we estimate in 3.6.10 are unlikely to be entirely driven by those observations nearest to the 250,000 MSA population cutoff.

To further examine the robustness of our estimates, we run a series of placebo test RDs in Figure 3.6.15 and 3.6.16. For in-network prices, we estimate an RD effect with a series of alternative cutoffs, ranging from 150,000 to 350,00 (a test in this style of a simulated -0.3 effect at the 250,000 cutoff can be seen in Appendix Figure C.1.4). For outpatient prices in Figure 3.6.15, we observe our estimated effect is generally largest at the actual cutoff of 250,000, while estimated effects are frequently smaller or insignificant at alternative cutoffs. For inpatient prices in Figure 3.6.16, we generally see effects that are insignificant or relatively small across the range of alternative cutoffs. These sets of results may help lend credence to our theory that these price effects are being driven predominately by a policy at the 250,000 MSA population urban cutoff, rather than some other nearby cutoff that may be biasing our estimates.

3.6 Conclusion

In this paper, we provide new evidence on the effects that insurer subsidies–even if short-lived–can have on insurance market characteristics. We leverage two critical policy changes to investigate this question. First, policy changes in 2001 and 2004 resulted in MA plans in urban counties receiving capitated payments from the government approximately 10.5% higher than those directed to plans in similar rural counties. Second, starting in 2012, the Affordable Care Act gradually phased out much of this benchmark discrepancy between urban and rural counties.

Although these elevated payments began to phase out in 2012, substantial differences in MA enrollment, MA penetration rate, and the number of MA plans offered in urban counties persisted well after the phase-out was complete. The persistence of these effects indicates that high subsidies can have large impacts on MA enrollment and market structure, and that these impacts may persist at a similar level long after a large reduction in the level of the subsidy is implemented.

We also provide novel evidence that MA benchmark payments may influence transaction prices for medical care. The impact of these increased payments on negotiated medical prices is theoretically ambiguous, as several factors, like MA enrollment and MA plan entry, could influence bargaining outcomes. Using county-level price indices derived from HCCI claims data, we find persistent decreases in the prices for outpatient procedures associated with plans participating in counties whose initial benchmark payments were higher as a result of the urban cutoff. For most years, we find negligible effects of these higher benchmark payments in urban counties on the prices for inpatient stays. We find consistent results across care categories and alternative sets of price index controls. And, we see the largest effects of the payments on prices in counties that have highly concentrated insurer markets. Like our analyses of enrollment, we find that these price effects are persistent across years.

Our price index spans the years 2012 through 2016, during which the differential in benchmark payments directed to MA plans in urban counties was largely phased out, while significant differences in MA enrollment in urban counties remained relatively stable. Additionally, in our analyses of entry, exit, and market concentration, we find that while more plans and insurers participated in these urban markets, enrollment outpaced entry, and we see only minor effects on market concentration. Given our findings of persistent negative price effects for outpatient procedures across all years, these results underscore the relative importance of the enrollment effect over the incomplete passthrough and entry effects. The incomplete passthrough effect should decrease as the differences in benchmarks fade and entry effects are largely absent across all years. However, enrollment remains persistently higher and outpatient prices remain persistently lower in urban areas across all years, indicative of the strong effects the number of covered lives may plan in insurer-provider network and price negotiations.

In this paper, we show that insurer subsidies can have large impacts on MA market structure. These subsidies and their market impacts, particularly higher levels of MA enrollment, can in turn have large impacts on negotiated prices for MA care, with heterogeneous effects seen across different types of care. These results help to underscore the importance of bargaining in MA markets, with the potential present for large decreases in the prices for medical care. However, the question remains of who benefits most from these subsidy, enrollment, and price changes. Do patients receive less expensive care, or are insurers "doubly subsidized," through both government subsidies and lower prices paid to providers? Future work could examine this issue, or explicitly model the bargaining problem between insurers and providers in an attempt to more accurately measure the relative importance of higher subsidies, additional market entry, and increases in the number of covered lives.

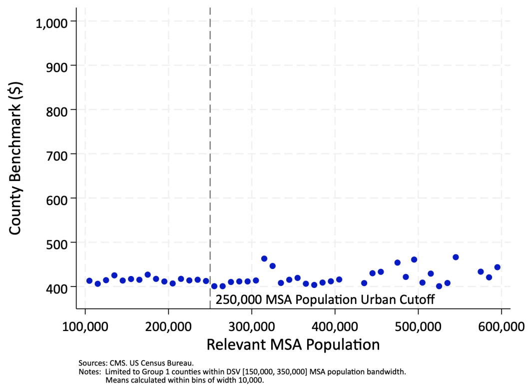
Table 3.6.1: RD Estimates of Urban Floor Cutoff on Price Indices, Outpatient Procedures

All Outpatient Procedures, 2012	-.283	-.34	-.339	-.388
	.082	.072	.102	.092
Bw	[150000,350000]	[97390.07,3642556]	[150000,350000]	[95022.53,2377744]
N (Left, Right of bw)	[201,174]	[93,352]	[201,174]	[86,332]
All Outpatient Procedures, 2013	-.212	-.256	-.258	-.315
	.093	.076	.112	.093
Bw	[150000,350000]	[105012.7,3795492]	[150000,350000]	[100217,2622931]
N (Left, Right of bw)	[199,170]	[100,349]	[199,170]	[95,335]
All Outpatient Procedures, 2014	-.269	-.353	-.325	-.428
	.1	.063	.125	.093
Bw	[150000,350000]	[99258.48,3874259]	[150000,350000]	[99199.06,3880324]
N (Left, Right of bw)	[194,167]	[90,341]	[194,167]	[90,341]
All Outpatient Procedures, 2015	-.314	-.292	-.361	-.38
	.082	.078	.103	.094
Bw	[150000,350000]	[121284,2731276]	[150000,350000]	[116728.5,3994780]
N (Left, Right of bw)	[194,171]	[120,333]	[194,171]	[119,346]
All Outpatient Procedures, 2016	-.194	-.185	-.225	-.252
	.067	.057	.073	.058
Bw	[150000,350000]	[114530.4,2569124]	[150000,350000]	[101345.5,3745957]
N (Left, Right of bw)	[201,175]	[122,338]	[201,175]	[97,352]
DSV Bandwidth	X	X		
Data Driven BW			X	X
RD Controls	X		X	

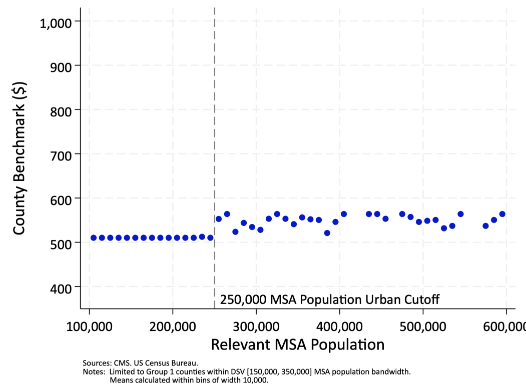
Sources: Authors' calculations from HCCI. CMS. US Census Bureau.
Notes: Limited to Group 1 counties.
SEs NN clustered at the relevant MSA. Bars represent 95% CIs.
Price Index Controls: Procedure Code, Log(Units), Demographics.
RD Controls: Log(1997 FFS Rate), Log(1997 MA Enrollment).

Figure 3.6.1: Mean County MA Benchmarks and Relevant MSA Populations

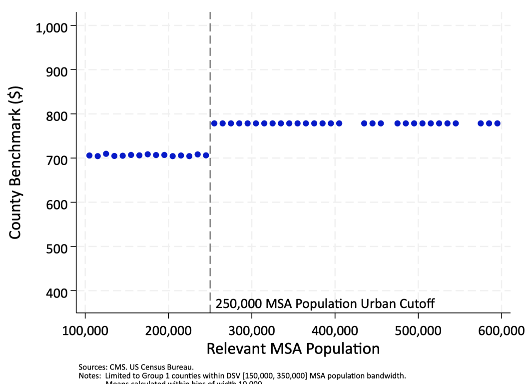
(a) 2000



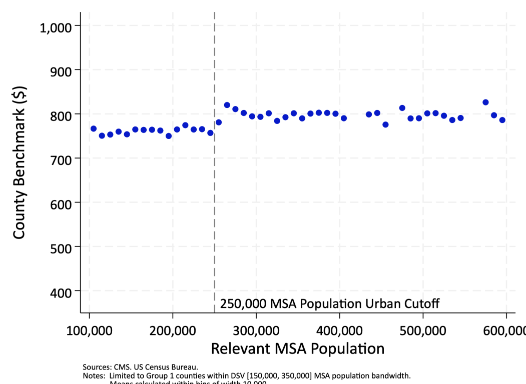
(b) 2003



(c) 2008



(d) 2014



(e) 2019

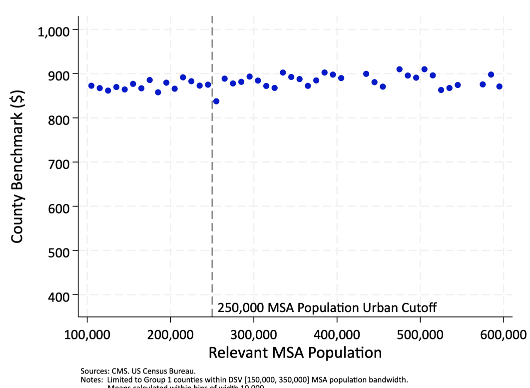
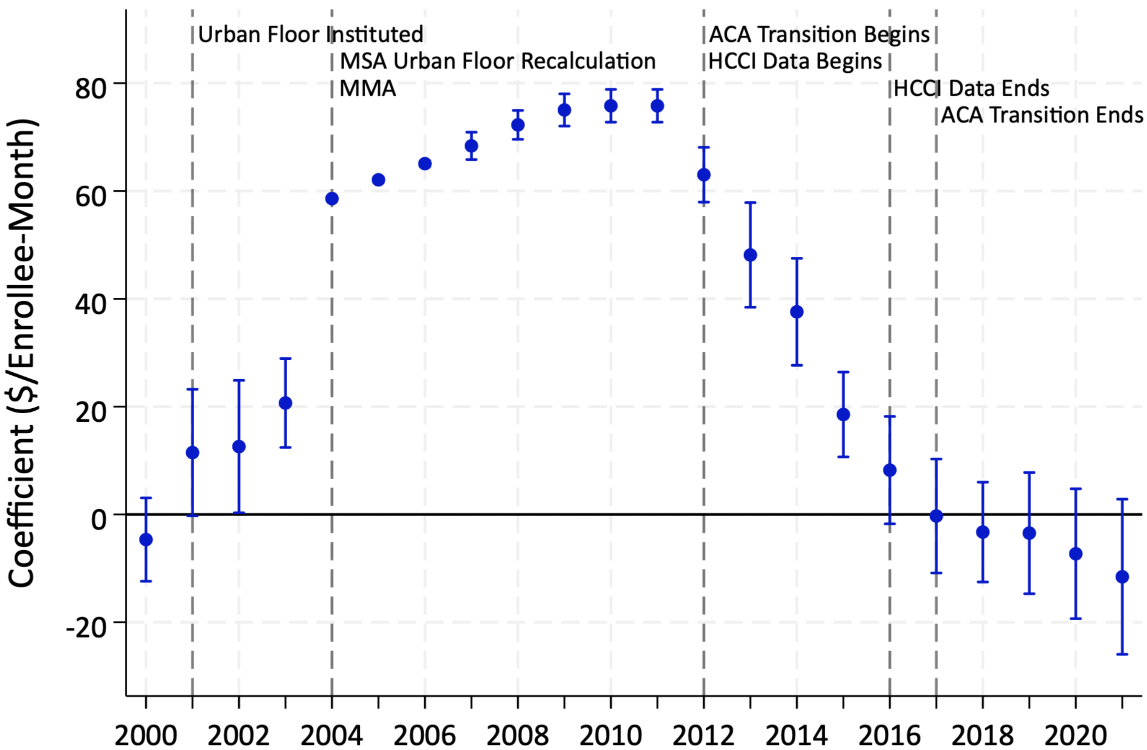


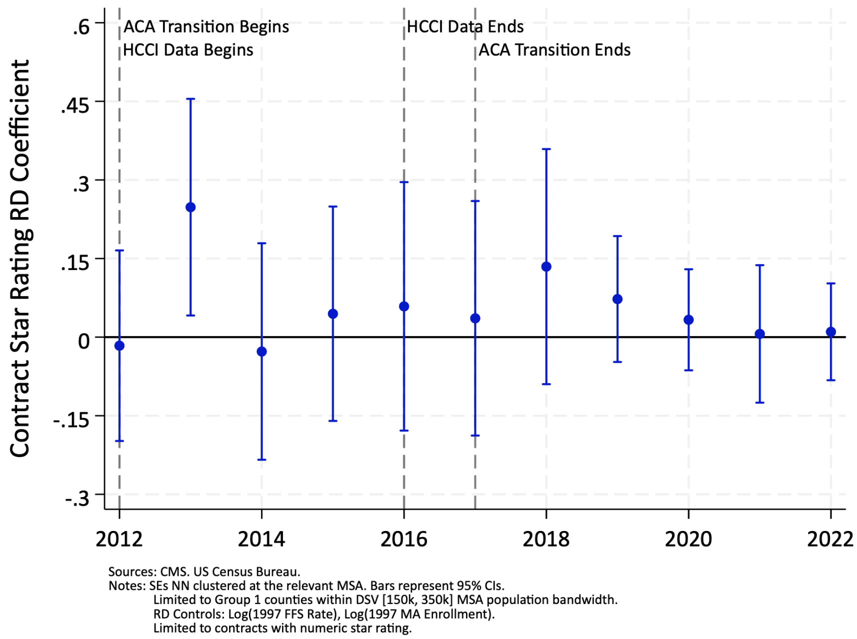
Figure 3.6.2: Estimates of Discontinuity in Benchmark Payments by Year



Sources: CMS. US Census Bureau.
Notes: Limited to Group 1 counties within DSV [150,000, 350,000] MSA population bandwidth.
SEs NN clustered at the relevant MSA. Bars represent 95% CIs.
RD Controls: Log(1997 MA Enrollment) and Log(1997 FFS Rate).

Figure 3.6.3: Estimates of Discontinuity in Benchmark Bonus Payments

(a) Average Contract Star Rating



(b) MA Bonus Benchmark

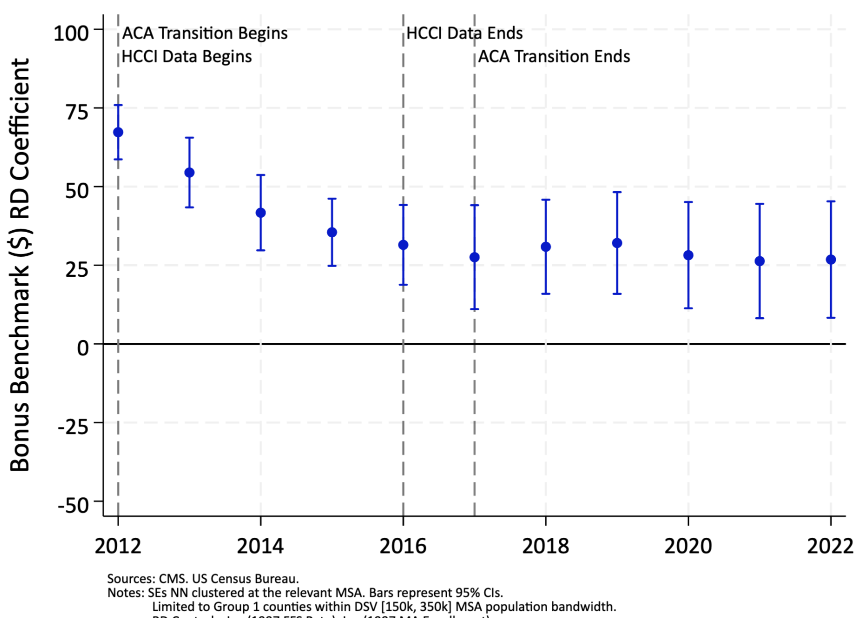
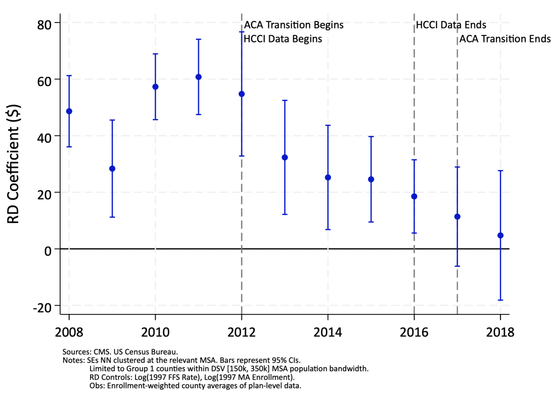
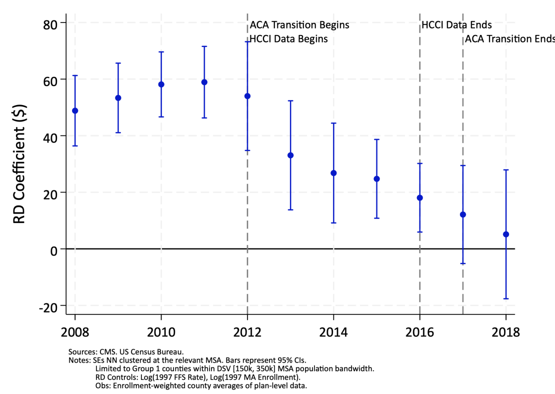


Figure 3.6.4: Estimates of Discontinuity in Bids and Rebates

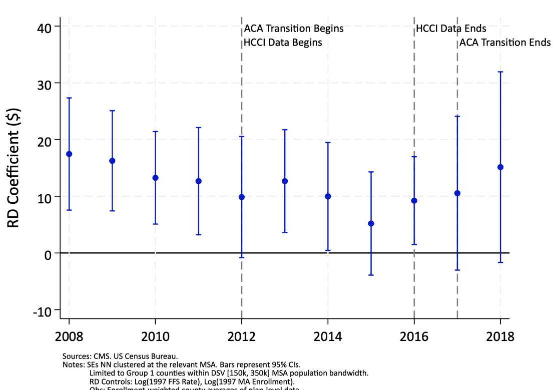
(a) Average MA Bid



(b) Average CMS Payment



(c) Average MA Rebate



(d) Average MA Payment + Rebate

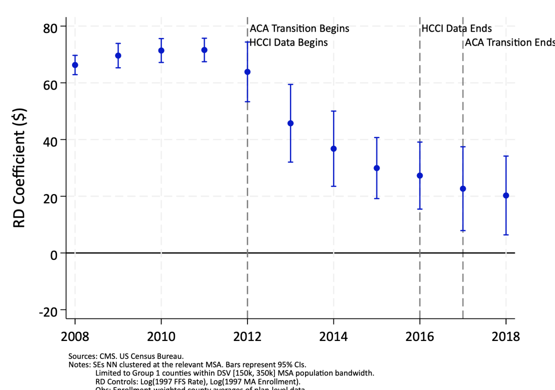
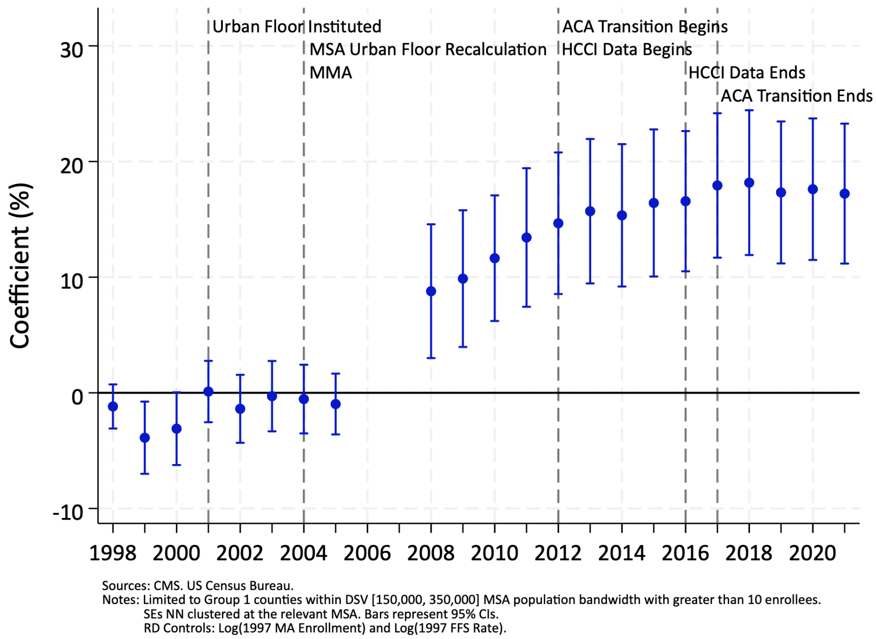


Figure 3.6.5: Estimates of Discontinuity in Medicare Advantage Penetration and Enrollment

(a) MA Penetration Rate



(b) Log(MA Enrollment)

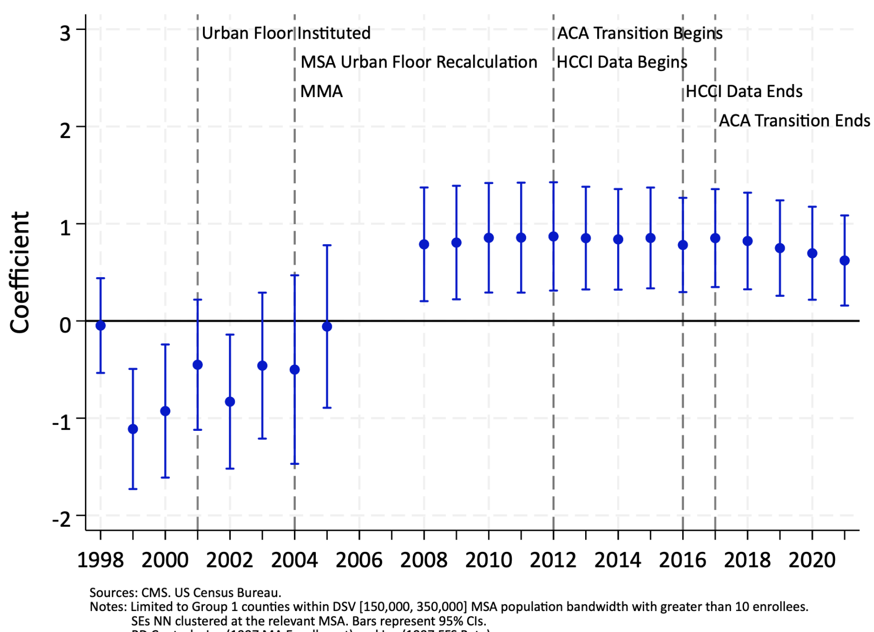
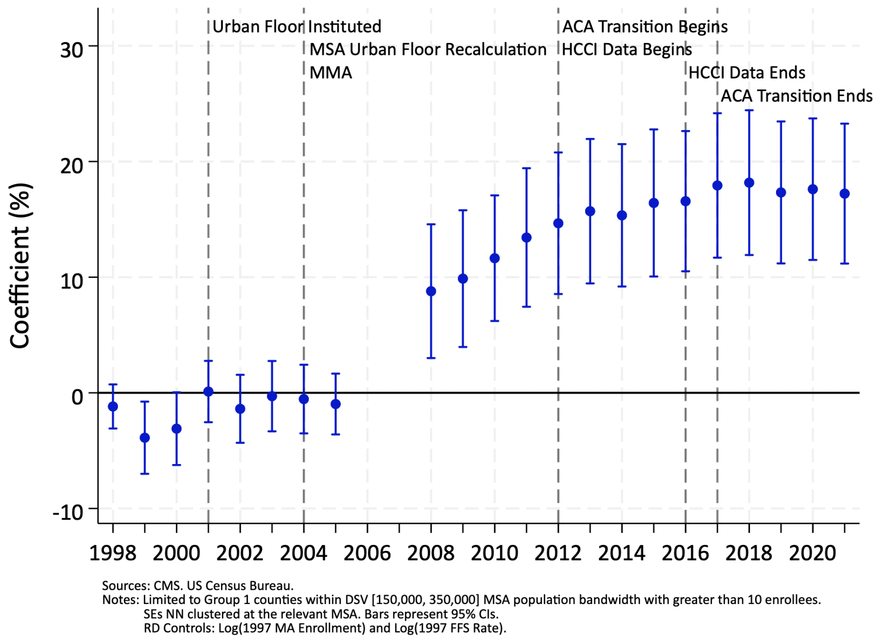


Figure 3.6.6: Estimates of Discontinuity in Medicare Advantage Penetration and Enrollment

(a) MA Penetration Rate



(b) Log(MA Enrollment)

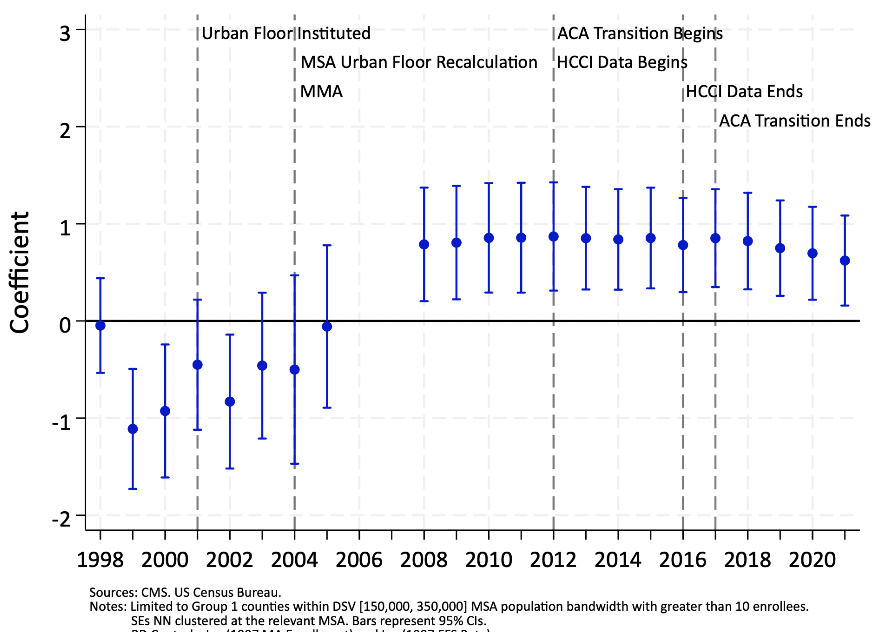
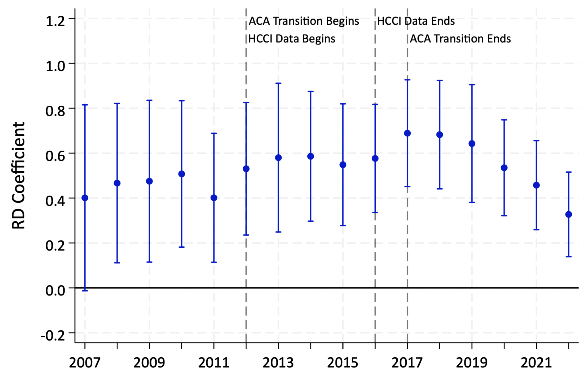


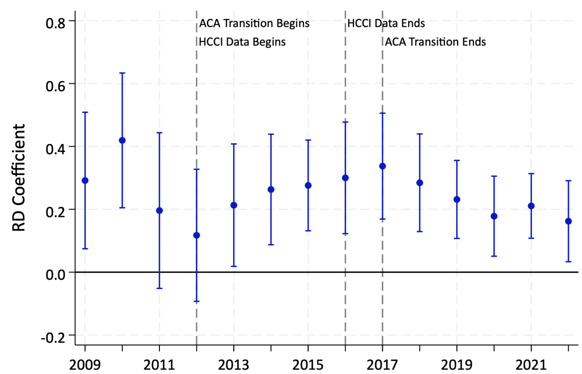
Figure 3.6.7: Estimates of Discontinuity in Medicare Advantage Plans and HHI

(a) Log(# of Plans)



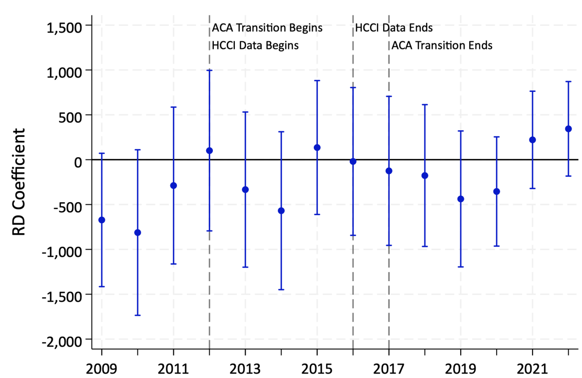
Sources: CMS. US Census Bureau.
Notes: Limited to Group 1 counties within DSV [150,000, 350,000] MSA population bandwidth and county-plans with greater than 10 enrollees.
SEs NN clustered at the relevant MSA. Bars represent 95% CIs.
RD Controls: Log(1997 MA Enrollment) and Log(1997 FFS Rate).

(b) Log(# of Parent Orgs)



Sources: CMS. US Census Bureau.
Notes: Limited to Group 1 counties within DSV [150,000, 350,000] MSA population bandwidth and county-plans with greater than 10 enrollees.
SEs NN clustered at the relevant MSA. Bars represent 95% CIs.
RD Controls: Log(1997 MA Enrollment) and Log(1997 FFS Rate).

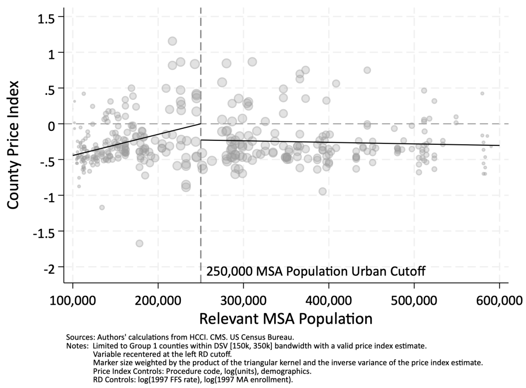
(c) HHI



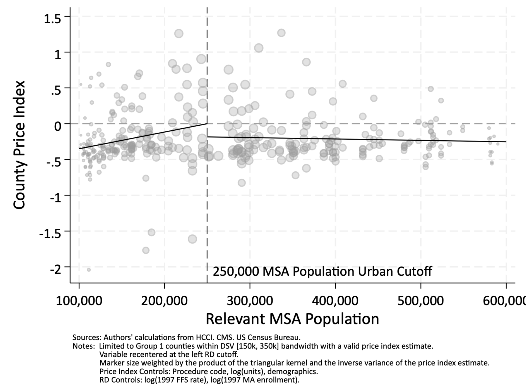
Sources: CMS. US Census Bureau.
Notes: Limited to Group 1 counties within DSV [150,000, 350,000] MSA population bandwidth and county-plans with greater than 10 enrollees.
SEs NN clustered at the relevant MSA. Bars represent 95% CIs.
RD Controls: Log(1997 MA Enrollment) and Log(1997 FFS Rate).

Figure 3.6.8: Estimates of Discontinuity in Outpatient Provider Prices

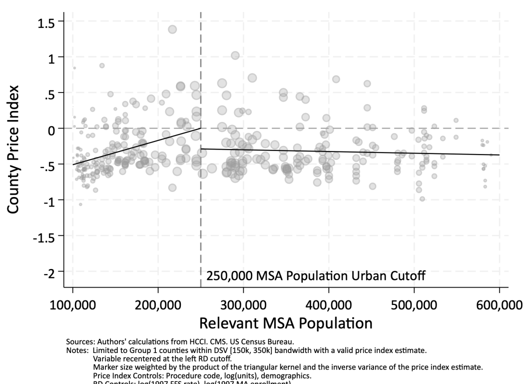
(a) 2012



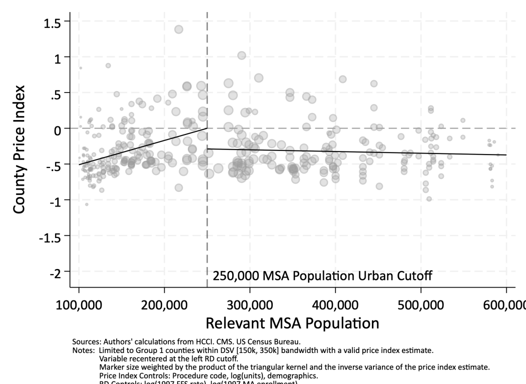
(b) 2013



(c) 2014



(d) 2015



(e) 2016

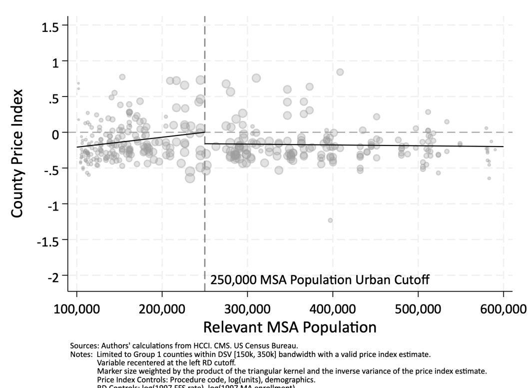
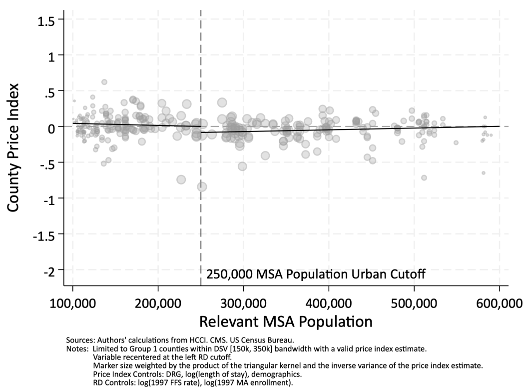
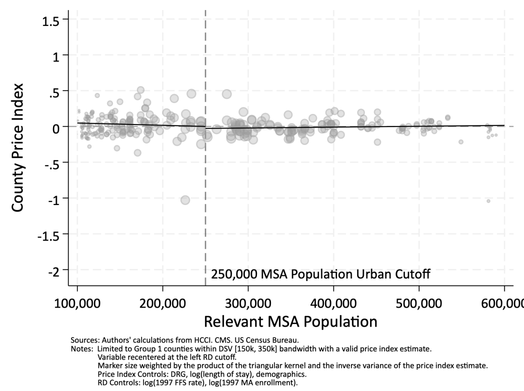


Figure 3.6.9: Estimates of Discontinuity in Inpatient Provider Prices

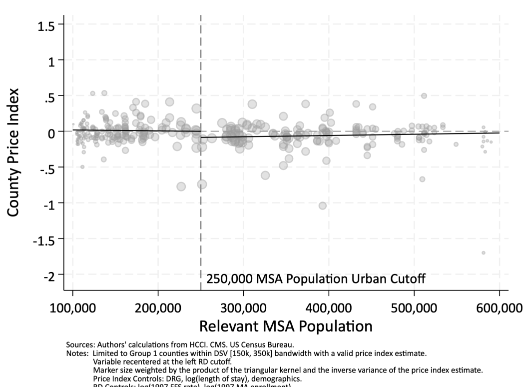
(a) 2012



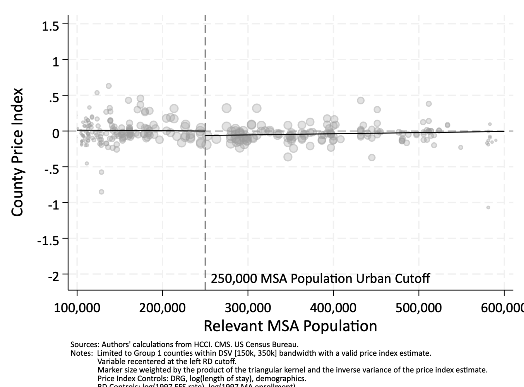
(b) 2013



(c) 2014



(d) 2015



(e) 2016

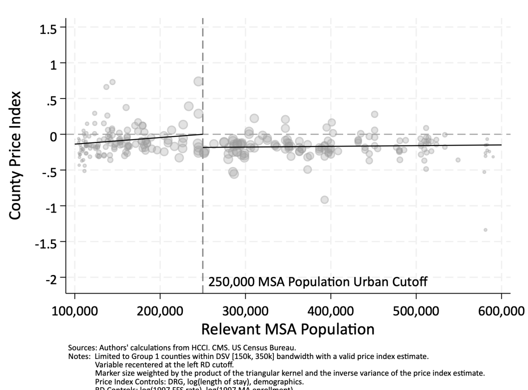
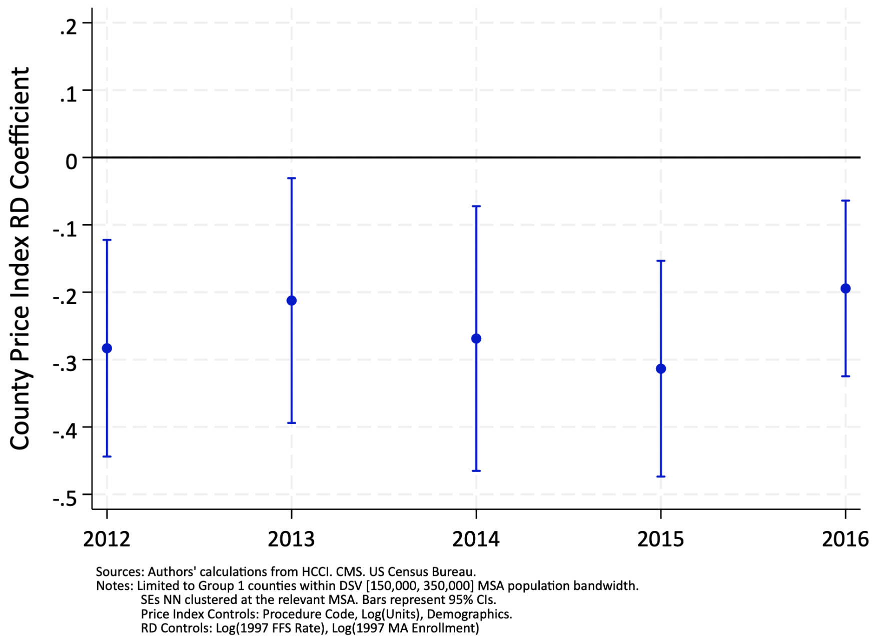


Figure 3.6.10: Estimates of Discontinuity in Provider Prices

(a) Outpatient Prices



(b) Inpatient Prices

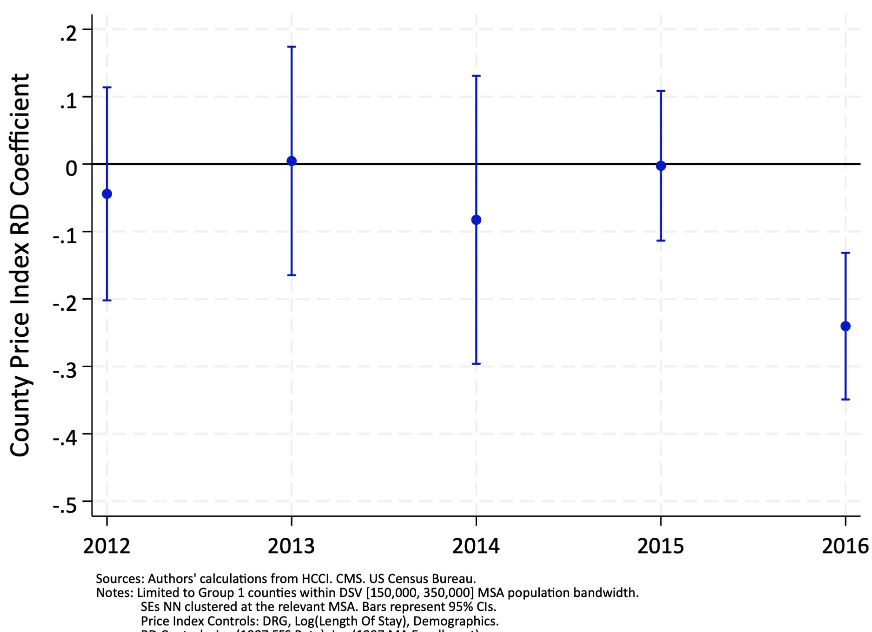
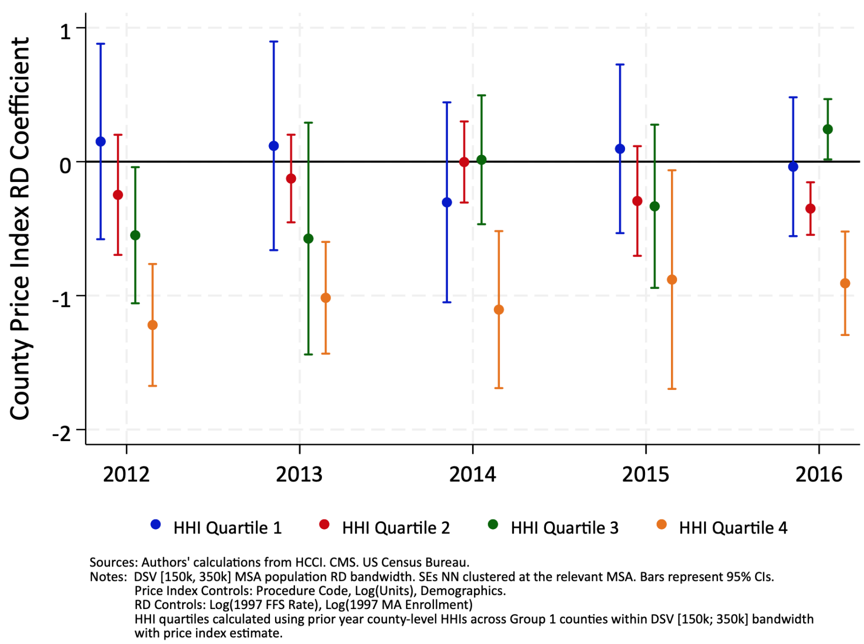


Figure 3.6.11: Estimates of Discontinuity in Provider Prices by HHI Quartile

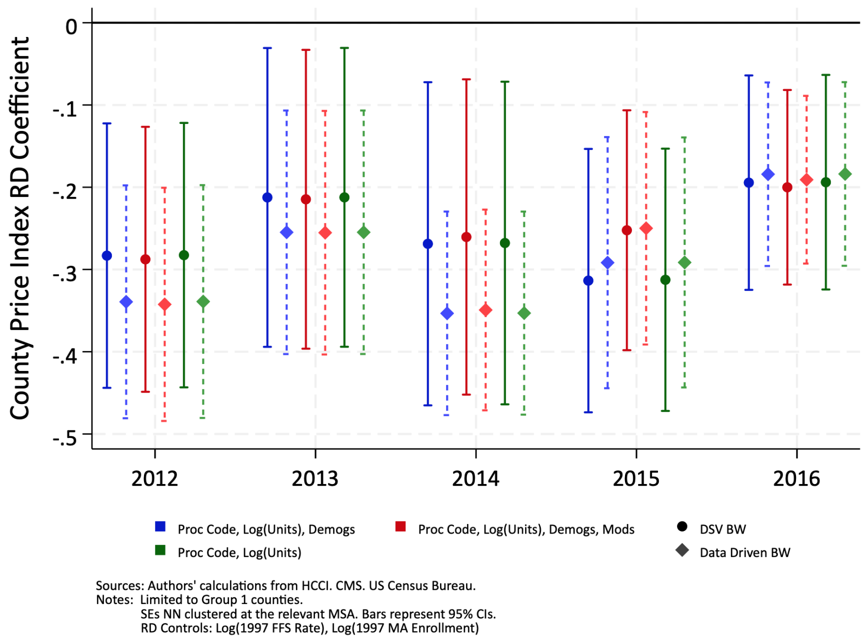
(a) Outpatient Prices



(b) Inpatient Prices

Figure 3.6.12: Estimates of Discontinuity in Provider Prices: Alternative Specifications

(a) Outpatient Prices



(b) Inpatient Prices

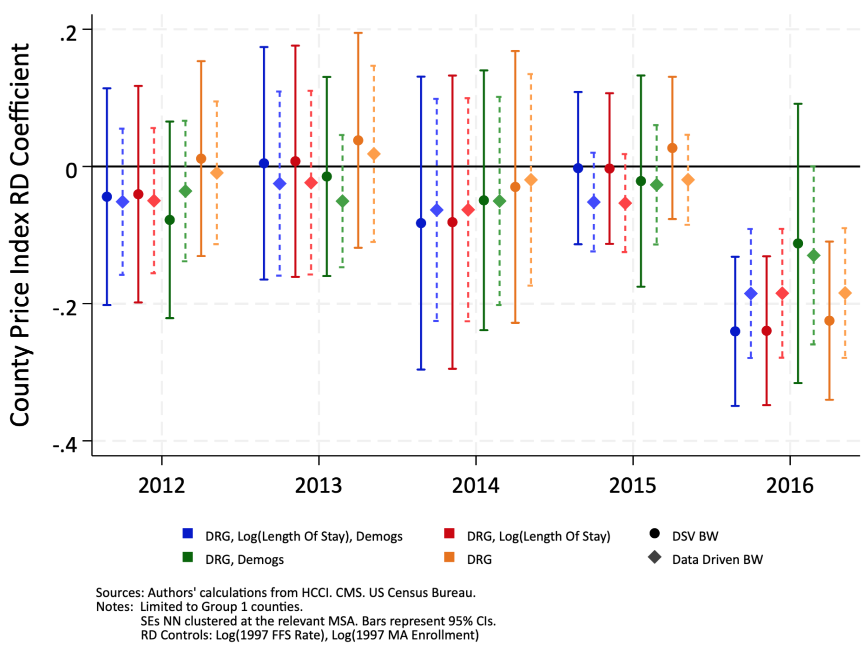
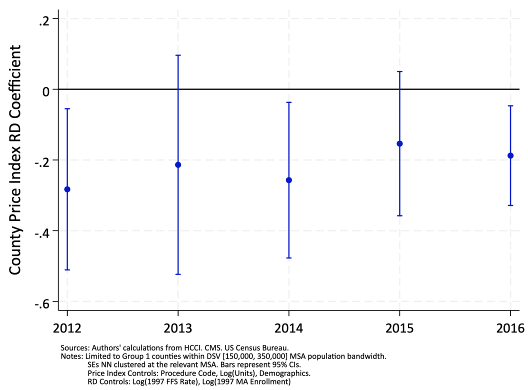
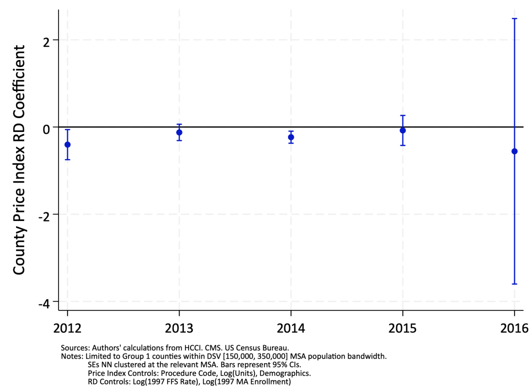


Figure 3.6.13: Estimates of Discontinuity in Provider Prices by Category (Part 1)

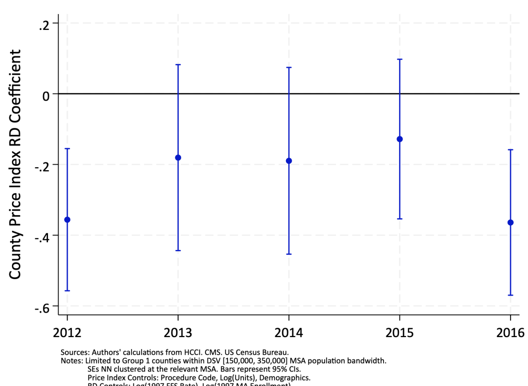
(a) Emergency Room (Outpatient)



(b) Ambulance Services (Outpatient)



(c) Radiology Services (Outpatient)



(d) Lab Pathology (Outpatient)

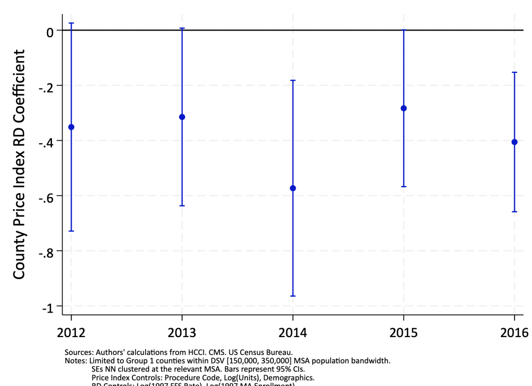
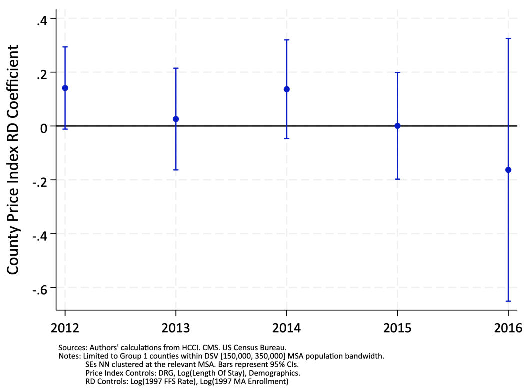
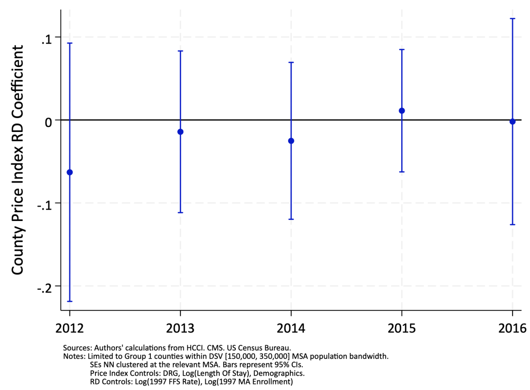


Figure 3.6.13: Estimates of Discontinuity in Provider Prices by Category (Part 2)

(e) Mental Health (Inpatient)



(f) Surgical (Inpatient)



(g) SNF (Inpatient)

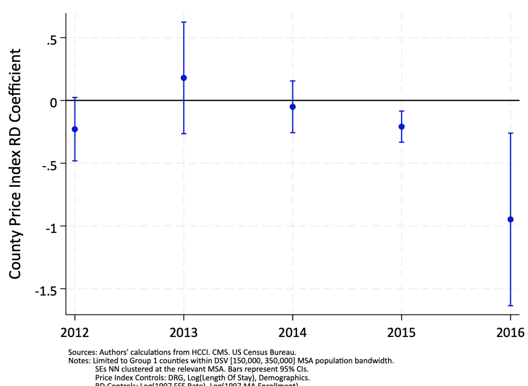
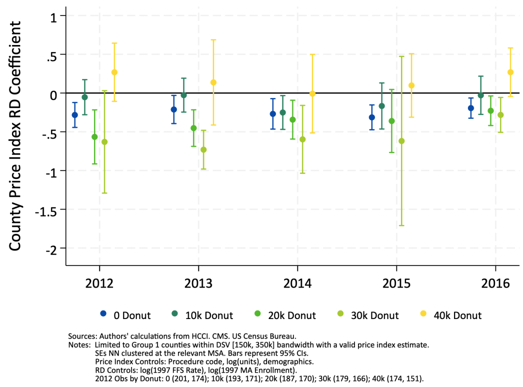
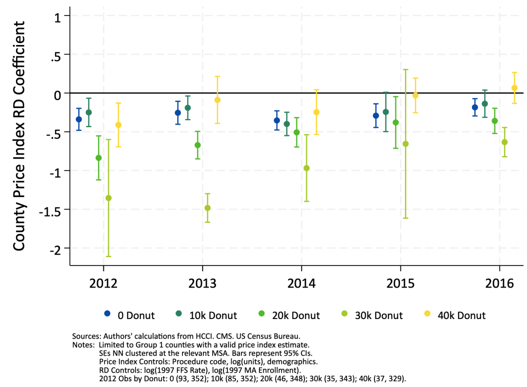


Figure 3.6.14: Donut Estimates of Discontinuity in In-Network Provider Prices

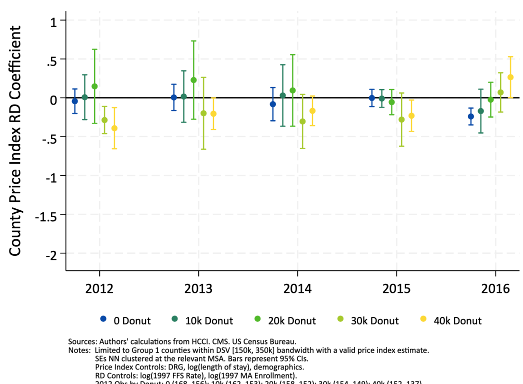
(a) Outpatient; Duggan et al. (2016) Bandwidth



(b) Outpatient; Data-Driven Bandwidth



(c) Inpatient; Duggan et al. (2016) Bandwidth



(d) Inpatient; Data-Driven Bandwidth

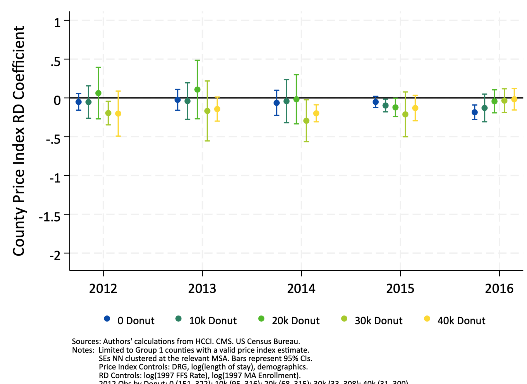
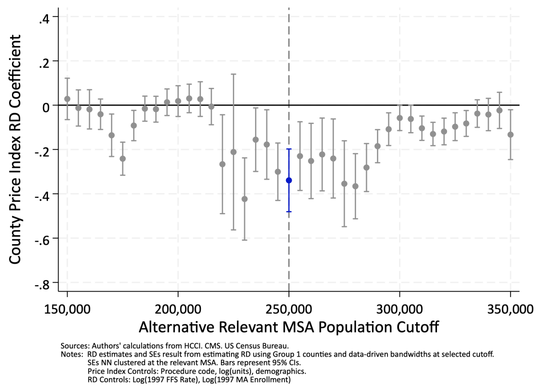
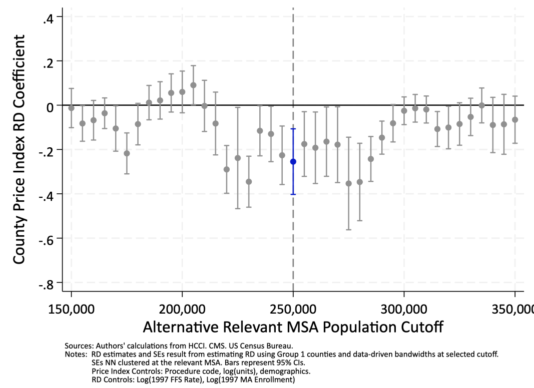


Figure 3.6.15: Placebo Estimates of Discontinuity in In-Network Outpatient Provider Prices

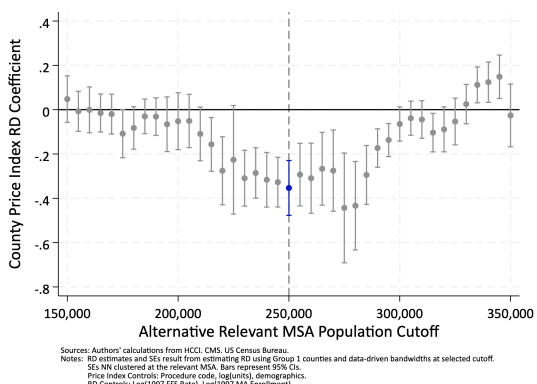
(a) 2012



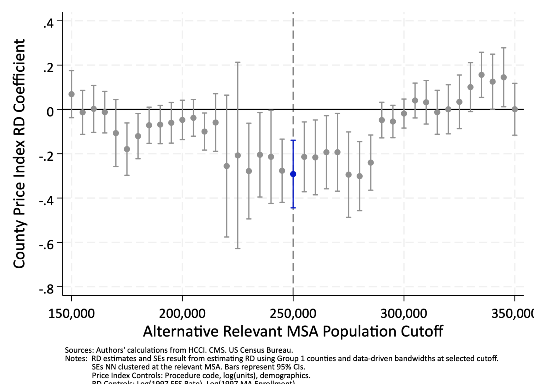
(b) 2013



(c) 2014



(d) 2015



(e) 2016

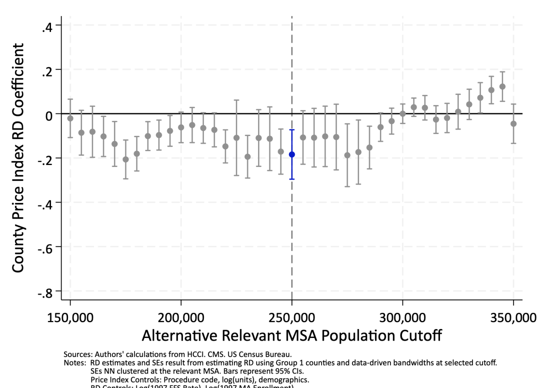
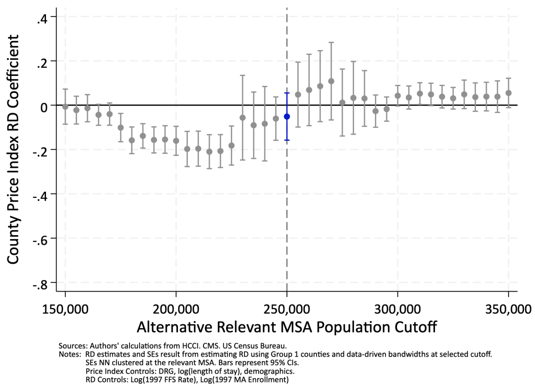
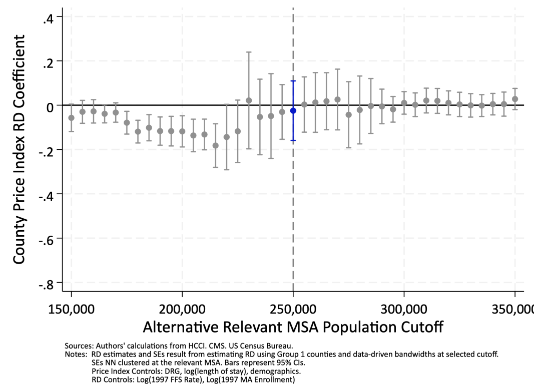


Figure 3.6.16: Placebo Estimates of Discontinuity in In-Network Inpatient Provider Prices

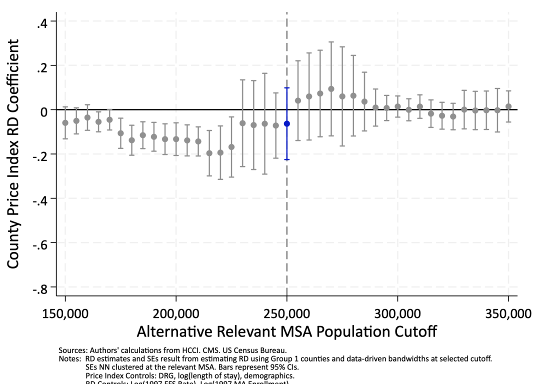
(a) 2012



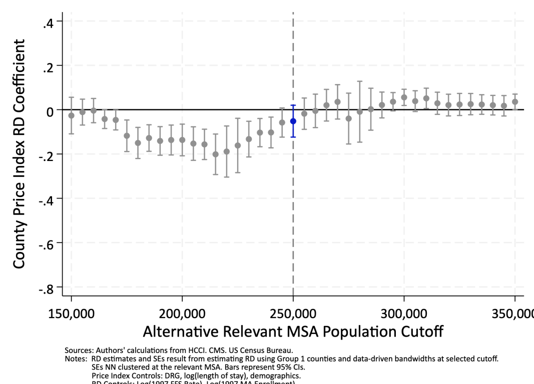
(b) 2013



(c) 2014



(d) 2015



(e) 2016

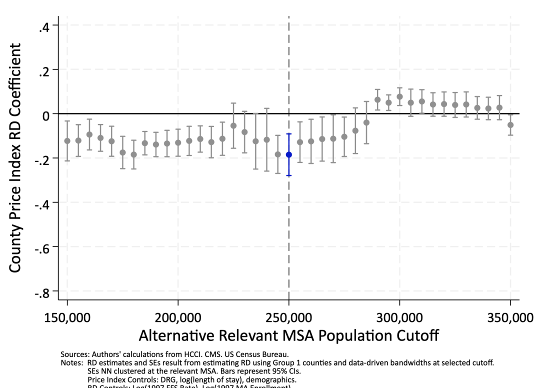


Table 3.6.2: RD Estimates of Urban Floor Cutoff on Price Indices, Inpatient Stays

All Inpatient Stays, 2012	-.044	-.051	-.048	-.068
	.081	.054	.083	.058
Bw	[150000,350000]	[142506.3,4394190]	[150000,350000]	[144352.1,4664443]
N (Left, Right of bw)	[168,156]	[151,322]	[168,156]	[157,336]
All Inpatient Stays, 2013	.005	-.025	.007	-.03
	.086	.068	.089	.07
Bw	[150000,350000]	[148466.6,4336270]	[150000,350000]	[149915.4,4287981]
N (Left, Right of bw)	[161,145]	[158,296]	[161,145]	[161,296]
All Inpatient Stays, 2014	-.083	-.063	-.087	-.084
	.109	.083	.116	.09
Bw	[150000,350000]	[142267.1,3929258]	[150000,350000]	[145786.2,3814576]
N (Left, Right of bw)	[164,151]	[149,309]	[164,151]	[161,309]
All Inpatient Stays, 2015	-.003	-.052	-.006	-.062
	.057	.037	.06	.038
Bw	[150000,350000]	[148283.9,3491561]	[150000,350000]	[150882.1,3591799]
N (Left, Right of bw)	[159,144]	[156,301]	[159,144]	[160,301]
All Inpatient Stays, 2016	-.24	-.185	-.238	-.187
	.055	.048	.055	.047
Bw	[150000,350000]	[144144.8,3765497]	[150000,350000]	[142303.9,3355111]
N (Left, Right of bw)	[135,140]	[126,275]	[135,140]	[122,273]
DSV Bandwidth	X	X		
Data Driven BW			X	X
RD Controls	X		X	

Sources: Authors' calculations from HCCI. CMS. US Census Bureau.
Notes: Limited to Group 1 counties.
SEs NN clustered at the relevant MSA. Bars represent 95% CIs.
Price Index Controls: DRG, Log(Length of Stay), Demographics.
RD Controls: Log(1997 FFS Rate), Log(1997 MA Enrollment).

APPENDIX A

Appendix to Chapter 1

A.1 Data

A.1.1 Additional Expedia Data Images

These images come from the competition website and the data summary from the IDCM 2013 presentation of the competition data.

Figure A.1.1: Kaggle Competition

Figure A.1.2: Expedia search can be preceded by Google or Bing search

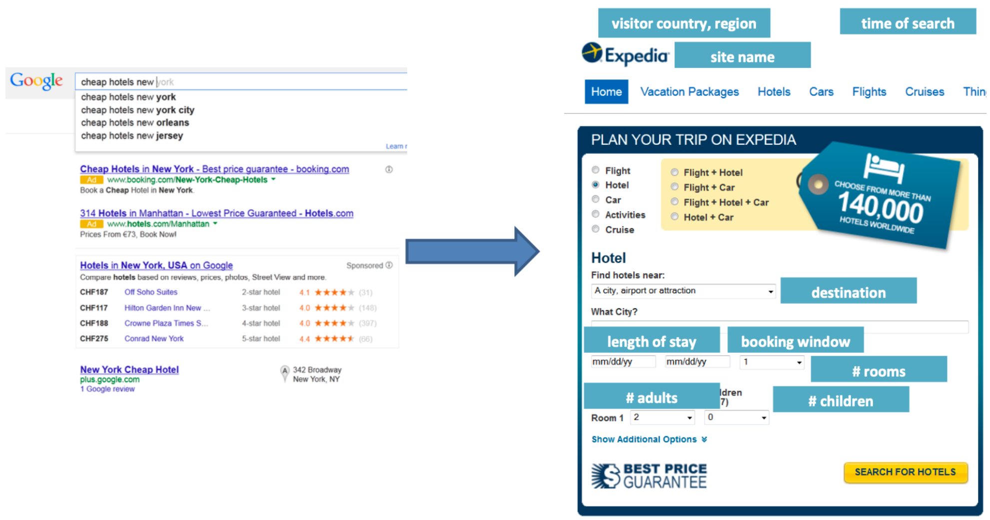


Figure A.1.3: Purchase Information

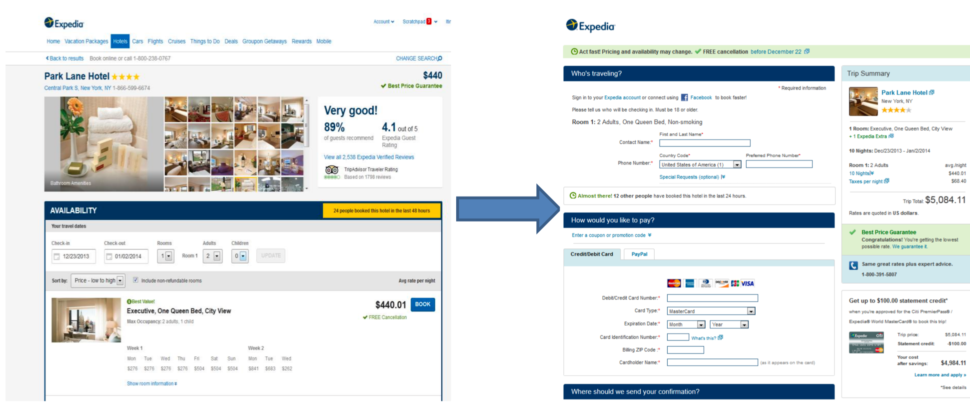
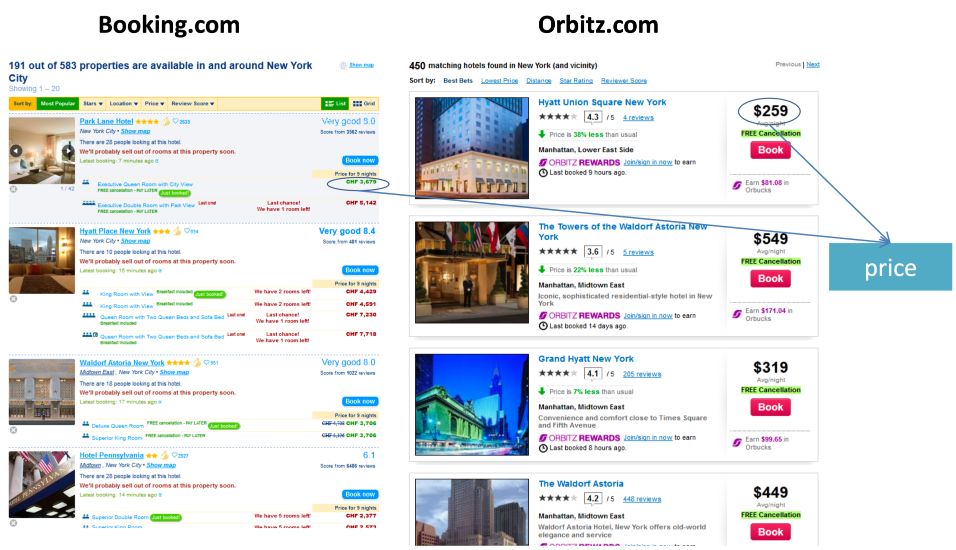


Figure A.1.4: Sometimes Expedia tracks prices and availability on other OTAs



A.1.2 Data Processing Details

This section highlights some of the key data processing steps. The Expedia data were released for a data science competition and are well-suited for training recommendation systems. There are, however, a few limitations that present difficulties in conducting the type of demand estimation and counterfactual analysis in this paper. This section highlights those challenges and the approaches that I use to address them. More details on the data processing approach are included in the appendix.

A.1.2.1 Market Definitions via K-means Clustering

In this section, I describe how I define markets from the deidentified data using a data-driven approach. The data include ID variables for properties, countries, and search terms but lack specific keys. For example, while an identifier might indicate "search term 52," there is no direct link to a specific term such as "Manhattan, NY." Multiple search terms could correspond to the same underlying market. As this paper evaluates a supply-side problem of hotel pricing behavior, it is crucial to ensure that I do not mistakenly exclude observations from a significant portion of the market. I define markets using K-means clustering, an unsupervised machine-learning technique. This clustering procedure matches search terms based on the similarity of their results. In plain terms, this procedure aims to match search terms that produce a common set of hotels in the displayed results. Further details on the k-means clustering procedure are included in the appendix.

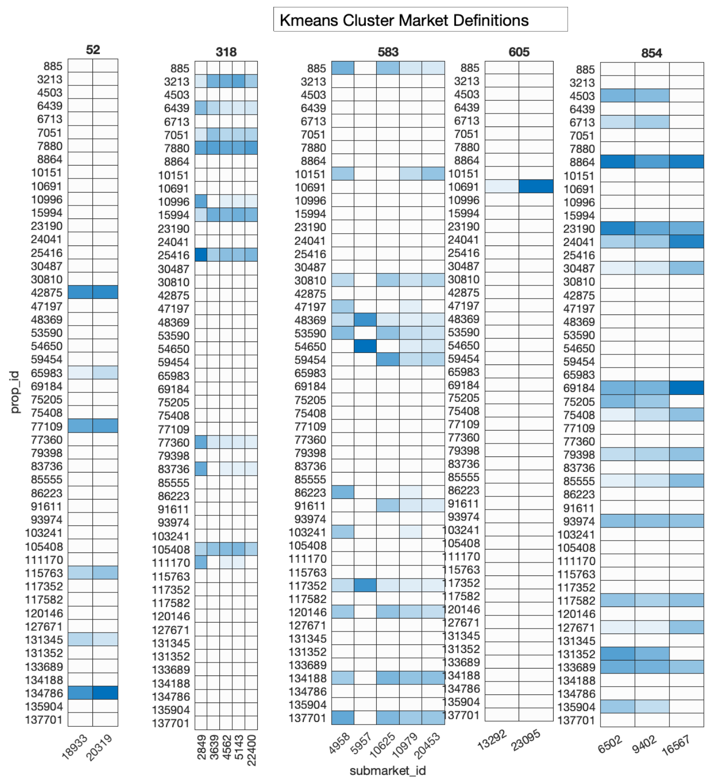
Figure A.1.5 displays cluster definitions for five markets. Each column stands for a search term, while the rows denote hotels. The shade of a hotel search term cell reflects the frequency with which that hotel appears in searches, weighted by position.

A.1.2.2 Final Transaction Price Prediction

A limitation of this dataset is that it records final transaction prices only when there is a purchase. When a result for a hotel is clicked but no purchase is made, the consumer may still discern the final transaction price, but this price is omitted from the data. Transaction prices are important for two reasons. First, they influence consumer search and purchase decisions. A consumer might, through a click, learn the final price, which informs their next search or purchase decisions. Second, for an accurate measure of consumer welfare, the final prices are essential, as these represent the actual expenditures by consumers.

To address this missing data issue, I impute the percent difference between the headline price and the final per-night transaction prices using the median hotel stay length. I use hotel-length of stay median, to impute the percent difference between headline prices and final

Figure A.1.5: Market Definitions by Search Term Clusters

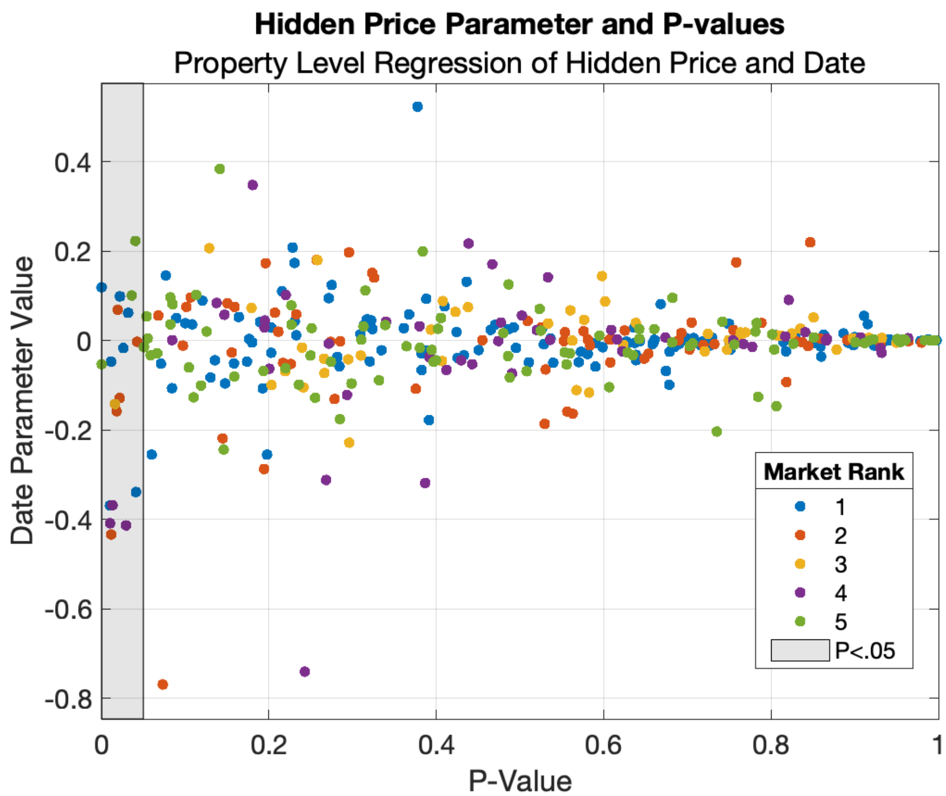


Notes: visualization includes subset of search terms matched to top 5 markets. The prop_id refers to hotels. The figure includes a subset of hotels that appears in on of the top markets. The k–means clustering procedure has one dimension for each hotel that appears in the data in at least five searches.

transaction prices. I use the hotel median for hotels with a limited number of transactions, while hotels with fewer than 3 observations are assigned the market-length of stay median. One potential concern with this methodology is that hotels could have modified their concealed pricing strategy during the study. To evaluate this concern I run a regression for each hotel in the top-5 markets with hidden price as the left hand side variable, and date as the right hand side variable. If many firms changed hidden pricing strategy during my period of study, I would expect many significant coefficients on date. Figure A.1.6, plots the results of this robustness check.

In Figure A.1.6, each point represents the property-specific coefficient on date from each regressions. The parameter value is plotted vertically, and the p-value is plotted horizontally. The shaded regions covers marks the estimates that are significant on the 5% level. f many firms changed hidden pricing strategy during my period of study I would expect to see

Figure A.1.6: Date parameter and P-value of hotel-level regressions on hidden price



bunching in the statistically significant region of the figure. Instead, roughly 5% of points are significant, which is in line with what one would expect to see just by chance.

A.1.2.3 Click Order Prediction

For each consumer-query, I observe which items each consumer clicked and which they purchased, but I do not know the order of clicks. For the naturally ordered data in the top five markets (by booking revenue), 92.5% of the consumer-sessions have only one click. I use a linear prediction model of clicks, where the click likelihood depends on the visible price, price interacted with promotions, the non-date visible product features, and a spline of slots. I then impute the order based on predicted click likelihood.

An alternative approach is to assume that the clicks happen in order of slot. However, this could overestimate position effects since this would force parameter estimates based on data where higher-ranked products are always clicked before lower-ranked ones. Another approach is to adjust the likelihood function. Alternative methodologies might assume that clicks occur in the sequence of the product slot.

A.1.3 Sample Selection

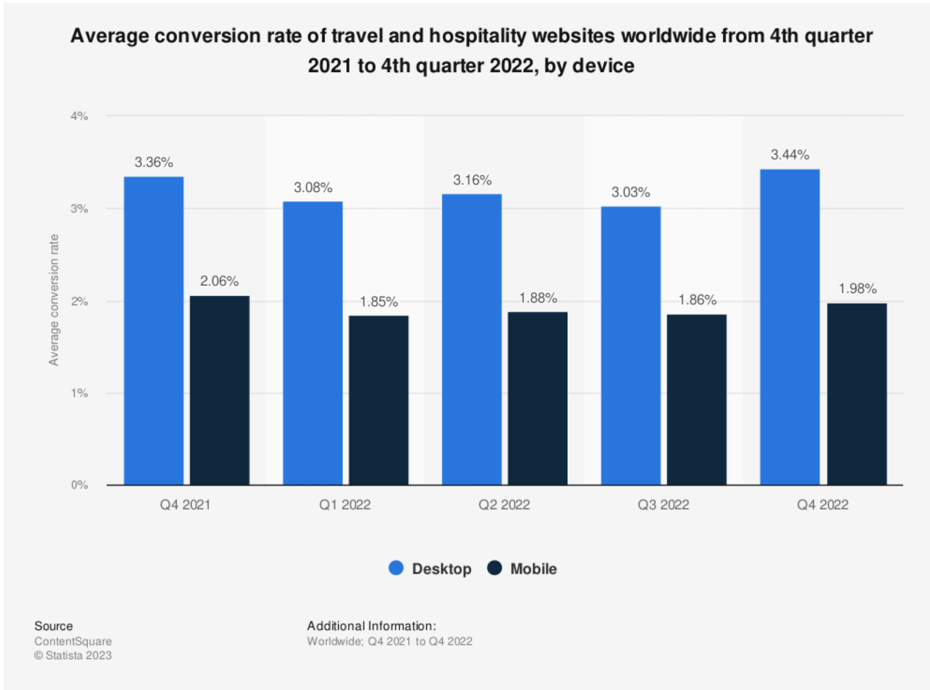
These data were initially intended for a data science competition aimed at developing recommendation systems. In that context, consumer-sessions that include a purchase are more valuable than consumer-sessions that have only clicks, and consumer-sessions with clicks tend to be more valuable than consumer-sessions without them. It is also important to note that e-commerce platforms often consider conversion rates proprietary information.

Three issues arise from the competition's data sampling method: 1) selection on clicks, 2) oversampling of transactions, and 3) ambiguity in the sample size. The following subsections detail each of these concerns and my approach to addressing them.

A.1.3.1 Selection on Clicks

The data include only consumer-sessions for which the consumer clicked at least one product. In demand estimation, I address this issue by using conditional likelihoods, conditioning each target consumer's joint likelihoods of observed clicks and purchases on the likelihood of clicking at least one product. On the supply side, I address this issue by using my parameter estimates from demand to reweight observations. The estimation section of this paper offers a more detailed discussion of these adjustments.

Figure A.1.7: Hospitality Average Conversion Rates



A.1.3.2 Oversampling of Transactions

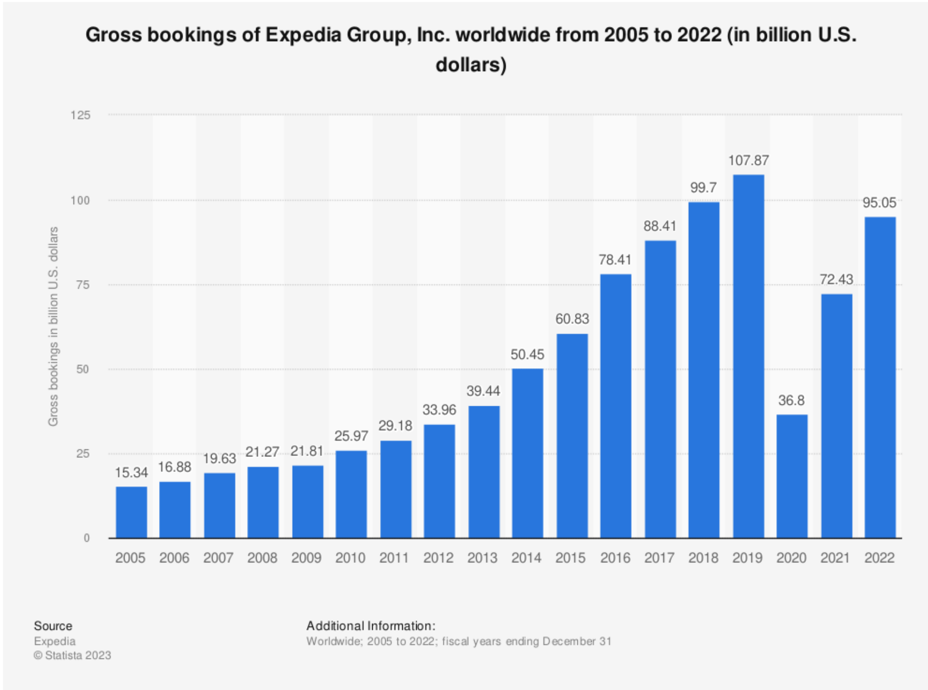
In the original competition data, observations are sampled randomly but at different rates depending on whether the observation includes a purchase or only clicks. Consequently, this means oversampling of observations in which the consumer made a purchase relative to those in which the consumer clicked but did not purchase anything (chose the outside option).

It is important to address this issue as I need to estimate the population distribution of preferences and capture substitution patterns in my counterfactual simulations.

Again, the true underlying conversion rates are confidential, so here, the task is to choose sensible parameters. To select the structural weights, I look to other sources for reasonable values. First, I need the percentage of searches with at least one click. I chose 30%—which is within the range of numbers reported in De los Santos and Koulayev (2016), which finds that 33% of searches resulted in a click for an OTA in Chicago, and Ursu (2018), which finds a 23% click rate for an OTA in Manhattan. Next, I need a conversion rate. I choose 5%, meaning that 5% of searches end in a transaction.[1] These numbers, together, imply a conversion rate conditional on at least one click of 16.67%. I then apply inverse sampling weights to the log-likelihoods during demand estimation to align with this conversion rate.

[1] 2021-2022 statistic on travel and hospitality websites range from 1.8-3.5%. Ursu (2018) notes a 1% rate in supplementary data.

Figure A.1.8: Reported gross bookings of Expedia Group, 2005-2022



source: https://www.statista.com/statistics/269386/gross-bookings-of-expedia/

While these weights might not capture the full picture, the proprietary nature of exact conversion rates necessitates imposing a reasonable structural assumption instead of attempting to recover exact conversion rates. An important next step of this project is to conduct a sensitivity analysis in line with Andrews et al.'s (2020) guidelines on transparency in structural modeling. In the exercise, I will assess the robustness of the primary findings by rerunning the main analysis under different assumed conversion rates.

A.1.3.3 Sample Rates

The last issue relates to sample size. The data's scope does not cover all Expedia searches and transactions. However, the granularity of the dataset, which includes gross booking revenue combined with Expedia's publicly disclosed revenue figures, permits broader extrapolation.

A.2 Demand Notes

This section provides additional detail on the features of the demand model.

I develop a novel optimal sequential search demand model, based on Weitzman (1979), and estimation strategy. To measure consumer welfare and have realistic substitution patterns in counterfactuals, my demand model accounts for important aspects of platform design, including four innovations over standard search models. The model can be applied more broadly to other online or offline settings where consumers engage in costly search and where the researcher observes search and purchase decisions. I develop a novel optimal sequential search demand model, based on Weitzman (1979), and estimation strategy. To measure consumer welfare and have realistic substitution patterns in counterfactuals, my demand model accounts for important aspects of platform design, including four innovations over standard search models. The model can be applied more broadly to other online or offline settings where consumers engage in costly search and where the researcher observes search and purchase decisions.

First, to account for feature emphasis, the demand model allows for both visible and hidden product features. This means that consumers know some of the product features prior to searching and learn about other product features after searching. In the context of Expedia, consumers know the product features that appear on the landing page and can learn the remaining product features by clicking through to the product-specific page. The standard search model assumption is that consumers have full information on product features and search only over an independent and identically distributed match quality term. This part of the model complements recent work by Compiani et al. (2021).

The second innovation relaxes the standard assumption that consumers learn the match term from search. Some recent work has introduced a visible and hidden term (Ursu et al., 2023; Morozov et al., 2021; Morozov, 2023). This paper builds on that by introducing a data-driven approach based on the variance structure of the match quality term used in nested logit established by Cardell (1997) and recent advances by Galichon (2022). With this data-driven approach, the model subsumes both the full-information demand model and the traditional search model.

The third innovation relates to the mechanisms underlying position effects. The standard approach in the empirical literature is to impose a structural assumption that position impacts demand only through search cost. I allow product ranking to influence demand both through search cost and through rational expectations. In terms of rational expectations, prior to a search, consumers form beliefs on the hidden product features conditional on position on the page. As a robustness check, I test this structural assumption against one where position impacts only search costs. To allow for these sophisticated beliefs, in estimation, I include a value function approximation in an inner loop. This setup introduces sufficient flexibility to test competing structural assumptions about consumer beliefs. For example, this also could allow for consumers' having higher-order beliefs about how the relationship between position on the page impacts the variance of hidden product features.

The fourth innovation takes advantage of a useful source of variation in the data. In my case, consumers arrive at the platform by searching for stays of different lengths. I focus on stays of one to four nights. Presumably, a consumer searching for a one-night stay and a consumer searching for a four-night stay face similar search costs. However, their returns to search are quite different. Since the consumer with the longer stay would consume more of the product and pay for each night, her returns to search are higher. This variation helps me separately identify consumer preferences from search costs, a common challenge in the search literature. This is consistent with other areas of the search literature in which consumers engage in more search when the returns to search are higher (Brown and Jeon, 2022). The best of my knowledge, this is the first paper to take advantage of variation in quantity in this way.

A.3 Model

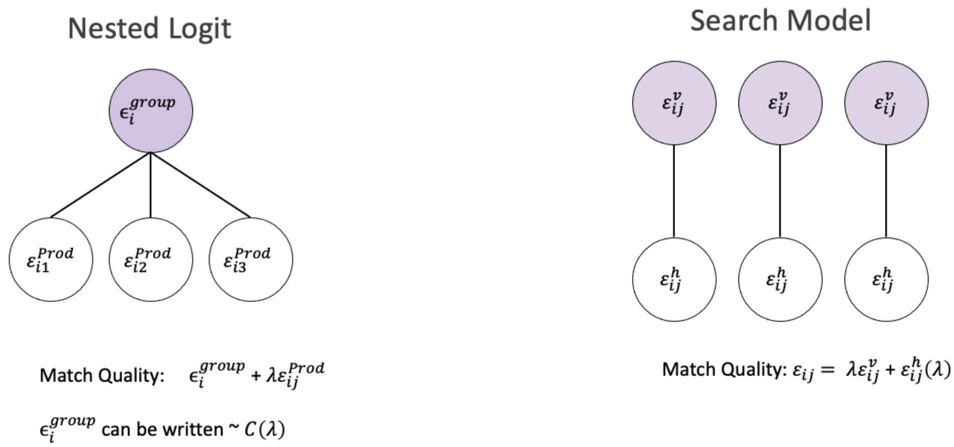
A.3.1 Structure of Match Quality Term

This section describes the details of the match quality term. In the demand model, I treat product features as either visible or hidden, with visible features appearing on the landing page. The match quality term follows a similar structure, with visible and hidden components, but with an added parameter λ that determines how much of the match quality term is known before search and how much is learned from search with along with the hidden product features. We can express the sum of the terms as:

$$\epsilon_{ijt} = \lambda\varepsilon^{v}_{ijt} + \varepsilon^{h}_{ijt}(\lambda) \tag{A.3.1}$$

where ϵ_{ijt} is consumer i's match quality for product j at time t and follows an i.i.d. type-1 extreme value distribution. ε^{v}_{ijt} is the match quality known before search and follows an i.i.d. type-1 extreme value distribution. It is multiplied by $\lambda \in (0, 1)$. ε^{h}_{ijt} follows a Cardell(λ) distribution , whose characteristic function depends on λ, and ε_{ijt} follows an i.i.d. type-1 extreme value distribution. As λ approaches 1, more of the match quality is visible, and the Cardell(λ) distribution collapses to zero. As λ approaches 0, more of the match quality is hidden, and the Cardell(λ) distribution approaches the type-1 extreme value distribution. The split error structure is similar to the nested logit, which includes nest-level and the item-level variance components of the error term. In my setup, both match quality terms are at the consumer–product–time level and are split based on the information available to the consumer. Identification of λ comes, in part, from the correlation between click decisions and product features.

Figure A.3.1: Split Error Term Structures: Nested Logit vs Search Model



Including the visible and hidden match quality term offers two advantages over reverting to the common assumption that consumers learn the entire error term from search. First, it overcomes one of the weaknesses of optimal sequential search models, where the search order is deterministic in product features. Second, the data-driven approach relaxes the assumption that the entire error is learned from search, allowing more model flexibility.[2] Recent papers, for example, Morozov et al. (2021), Morozov (2023), and Ursu et al. (2023) use normal distributions (double-probit) instead of extreme value (EV) and normalize one of the match quality terms. In contrast, I use the properties of the Cardell distribution Cardell (1997) and extreme distributions to create an EV-1 combined match quality term, which provides additional benefits. Since I normalize ϵ_{ijt} to be EV-1, the other utility parameters are scaled to this combined ϵ_{ijt} term instead of the hidden or visible component of match quality. This approach allows a straightforward interpretation of the utility and search cost parameters consistent with that under more conventional demand models. My setup also provides for within-simulation analytic expressions of choice and click likelihoods, which are necessary for the supply-side estimation and counterfactual simulations.[3]

A.3.1.1 Approximating the Cardell Distribution

Having discussed the split match quality term, I now turn to the challenges of taking draws from the Cardell distribution. Cardell (1997) proves the existence of the distribution. The nested logit, for example, used in Berry (1994), implicitly depends on the Cardell distribution

[2]In fact, this nests both the full-information demand ($\lambda = 1$) and sequential search ($\lambda = 0$) models.

[3]By within-simulation expression, I mean that I take simulated draws of all random parameters except ε_{ijt}^{v} and then use the property that ε_{ijt}^{v} is also EV-1 to construct choice and click likelihoods.

to create an analytic expression of choice probabilities and market shares and and identify a parameter λ from the diversion ratio. In contrast, my estimation strategy requires taking draws from the Cardell distribution, which is not straightforward since the distribution does not have a closed-form probability density function (PDF) or cumulative distribution function (CDF) and there is no ready to use software package to take these draws.

A recent advance by Galichon (2022) proves that the Cardell distribution with parameter λ is related to the positive stable distribution with parameter λ, specifically if $Z \sim P_{\text{stable}}(\lambda)$, where $P_{\text{stable}}(\lambda)$ is the stable distribution. With this property, the problem of drawing from the Cardell distribution reduces to sampling from a positive stable distribution and applying the correct transformation. In practice, drawing from the stable distribution is slow since each point is solved numerically. To avoid this slowdown, I create a fine grid of points in terms of λ and v, the probability value, numerically solved for the point in the inverse CDF of the Cardell distribution. Then, I create an interpolation object $f(\lambda, v)$.

A.4 Estimation

A.5 Demand Estimation and Results

A.5.0.1 Demand Estimation Procedure

Algorithm 1 summarizes the estimation procedure.

Algorithm 1 Demand Estimation Procedure

procedure PRE-ESTIMATION
Select Hyperparameters
Generate Splines
Sample Beliefs Data
Take Scrambled Halton Draws $D = 400$
Estimate Expected Hidden Features: $E[x^h|\Omega]$
Estimate Expected Price: $E[p_{ijt}|\Omega]$
end procedure
procedure MAXIMUM SIMULATED LIKELIHOOD
Initialize Parameters θ
while Not Converged **do**
Update Random Coefficients Using Draws and Cholesky Factor
Approximate Hidden-Match Quality $\varepsilon_{ijt}^{h[s]}(\lambda) = \text{ICDF}(\lambda, d_{ijt}^{[s]})$
Calculate Utilities and Search Costs
Calculate Expected Hidden Utility $E[\delta_{ijt}^{h[s]}|\Omega_{it}] = -e^{\rho_i^{[s]}} E[p_{ijt}|\Omega_{it}] + \beta_{ijt}^{h[s]} E[x_{ijt}^{h[s]}|\Omega_{it}]$
procedure RESERVATION UTILITY VALUE FUNCTION APPROXIMATION
Initialize Grid of State Variables
Solve for ζ at Each Point on Grid
Fit Spline Interpolation Object to ζ and Grid of State Variables
Predict Each $\zeta_{ijt}^{[s]}$ Using Interpolation Object
end procedure
procedure LOGIT SMOOTHING
Scale Per-Night Utility and Reservation Utility by Length of Stay
Apply Logit-Smoothing Parameter ω
end procedure
procedure LIKELIHOODS
Calculate Click, Purchase, and Joint Likelihoods
Calculate Any-Click Likelihood
Calculate Conditional Likelihood
Calculate Log-Likelihood
Apply Weights
end procedure
Calculate Weighted Log Likelihood
Calculate Gradient $\nabla_\theta \mathcal{L}(\theta)$ (Finite-Differences)
Update Parameters $\theta \leftarrow \theta + \alpha \nabla_\theta \mathcal{L}(\theta)$
Check Convergence Criteria
end while
return θ
end procedure

A.5.1 Search Cost and Slot Results

Figure A.5.1 plots the distribution of search cost (in utils) by position on the page.

Figure A.5.1: Search Cost Distribution

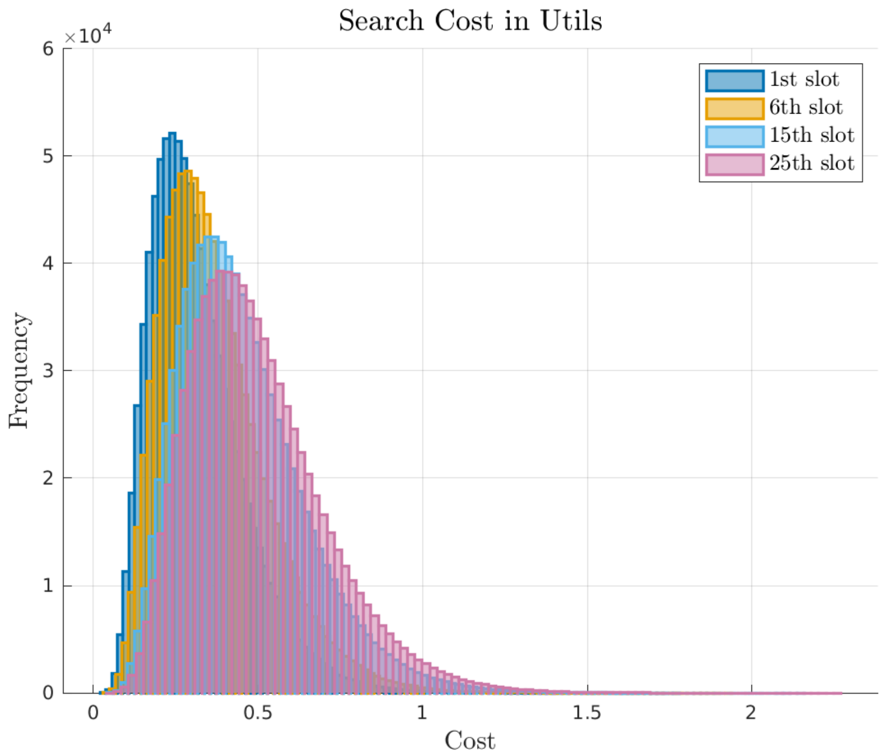


Table A.5.1 shows the implied median search cost, per-night utility, and reservation utility by slot and number of nights. Here, we see the dynamic where reservation utility increases with the length of stay.

Table A.5.1: Demand Results: Median Search Cost and Per-Night Utilities

Slot	**Search Cost ($)**	**Utility ($)**	**Reservation Utility ($)**			
			1 night	2 nights	3 nights	4 nights
1	99	97	563	871	995	1141
6	143	57	341	665	774	943
15	163	29	260	577	710	866
25	181	24	202	516	669	808

A.6 Supply Side Estimation and Results

A.6.1 Expected Quantity

While the demand estimation relies on logit-smoothing to estimate the utility and search cost parameters, once I have the parameter estimates, the properties of the split error structure allow me to estimate the choice, click, and selection likelihoods without smoothing. To calculate the expected quantity at the observed prices I sum choice weighted probabilities using both the demand and platform models.

Choice probability

Choi et al. (2018) show that under optimal sequential search, consumers choose the product with the highest $\min(u_{ijt}, r_{ijt})$. This is where the split-error distribution described in Section **??** becomes useful. Recall the visible component of match quality $\lambda\varepsilon^{v}_{ijt}$, where λ is a scale term estimated in the demand model and ε^{v}_{ijt} follows a type-1 extreme value distribution. These two properties let me construct an analytic expression for simulation level choice probabilities (without needing to use logit smoothing). To start, we need a term for the min of reservation utility and utility that excludes the visible error term (from now on called the *adjusted value*):

$$\mu^{[s]}_{ijt} = \frac{\min(u^{[s]}_{ijt}, r^{[s]}_{ijt}) - \lambda\varepsilon^{v[s]}_{ijt}}{\lambda} \tag{A.6.1}$$

$u^{[s]}_{ijt}$, $r^{[s]}_{ijt}$ share the same visible error term $\varepsilon^{v[s]}_{ijt}$, so I subtract to $\varepsilon^{v[s]}_{ijt}$ get the adjusted value. I also divide by λ to simplify the expression for choice probabilities. I make the same adjustments to the outside.

From here, we can construct the choice likelihoods using the demand model and observed rankings:

$$Pr(i \text{ choose } j|\theta) = \frac{1}{D}\sum_{s=1}^{D} \frac{\exp\left(\mu^{[s]}_{ijt}\right)}{\exp\left(\mu^{[s]}_{i0t}\right) + \sum_{k\in J_{it}} \exp\left(\mu^{[s]}_{ijt}\right)} \tag{A.6.2}$$

where θ are the parameter estimates and J_{it} is the set of all the hotels on the first page.

Observation Weights

I can similarly use the effective values, $\mu^{[s]}_{ijt}$ to calculate the observation weights. This also requires $\varrho^{[s]}_{ijt}$, which is the scaled reservation utility excluding the visible match quality term. The first step is to calculate the selection probability,$P^{\text{selection}}_{it}$, of each consumer-query.

$P_{it}^{\text{selection}}$ depends on the sampling weights (w^{in},w^{out}), the likelihood of clicking at least one product and making a purchase, and the likelihood of clicking at least one thing and choosing the outside option (no purchase).

$$P_{it}^{\text{selection}} = \frac{1}{D}\sum_{s=1}^{D}\left[\overbrace{\frac{1}{w^{\text{in}}}\left(\frac{\sum_{k\in J_{it}}\exp\left(\mu_{ijt}^{[s]}\right)}{\exp\left(\mu_{i0t}^{[s]}\right)+\sum_{k\in J_{it}}\exp\left(\mu_{ijt}^{[s]}\right)}\right)}^{\text{sampling likelihood from click and purchase}}\right.$$

$$+\left.\overbrace{\frac{1}{w^{\text{out}}}\left(\frac{\sum_{k\in J_{it}}\exp\left(\varrho_{ijt}^{[s]}\right)}{\exp\left(\mu_{i0t}^{[s]}\right)+\sum_{k\in J_{it}}\exp\left(\varrho_{ijt}^{[s]}\right)}-\frac{\sum_{k\in J_{it}}\exp\left(\mu_{ijt}^{[s]}\right)}{\exp\left(\mu_{i0t}^{[s]}\right)+\sum_{k\in J_{it}}\exp\left(\mu_{ijt}^{[s]}\right)}\right)}^{\text{sampling likelihood from click w/o purchase}}\right] \tag{A.6.3}$$

The observation weights are the inverse sampling likelihood from the demand model

$$\hat{w}_{it}^{obs} = \left(P_{it}^{\text{selection}}\right)^{-1} \tag{A.6.4}$$

Now we can express the expected quantity as

$$E[q_{jtt'}|\hat{\theta}_{\text{est}}] = \sum_{i\in I_{tt'}}\left[\hat{w}_{it}^{obs}\left(\frac{1}{D}\sum_{s=1}^{D}x_{it}^{nights}P_{it}^{[s]}\left(i\text{ chooses }j|\hat{\theta}^{[s]}\right)\right)\right] \tag{A.6.5}$$

where $E[q_{jtt'}|\hat{\theta}_{\text{est}}]$ is the expected quantity of product j, for stays in period t, from searches happening in period t'. $I_{tt'}$ denotes the set of consumers searching for stays in period t, from searches happening in period t'. Since the choice probabilities are based on the estimated population distribution of utility parameters $\hat{\theta}$, I multiply by the inverse of the selection likelihood, $\hat{w}_{it}^{obs}$. Note that when estimating elasticity and in counterfactual simulations, $\hat{w}_{it}^{obs}$ remains fixed, but $P_{it}^{[s]}\left(i\text{ chooses }j|\hat{\theta}^{[s]}\right)$ will change due to different prices and product rankings.

A.6.1.1 Own Price Elasticity

I calculate the own price elasticity via finite differences. The elasticity requires both the platform and demand model results. The elasticity requires the following steps outlined in algorithm 2

Algorithm 2 Own-Price Elasticity Procedure

1: Identify the target product j^* and subperiod
2: Increase j^*'s final and headline prices by $\epsilon\%$ ▷ Small + perturbation
3: Calculate ψ_{ij^*t} using the platform model ▷ Deterministic relevance score
4: Re-assign slots using ψ, β^{slot}, and $\varepsilon_{ijt}^{\text{rec}[s]}$
5: Update expectations of hidden features based on new slots
6: Recompute utilities, search costs, and reservations with new prices and slots
7: Derive new effective values $\mu_{ijt}^{[s]}$
8: Compute expected quantity, $q_{jtt'}^{+}$ from Eq. A.6.5 ▷ Subperiod demand
9: Repeat steps 1-6, instead decreasing price by $\epsilon\%$ ▷ Small - perturbation for $q_{jtt'}^{-}$
10: Calculate arc elasticity for j^* using $q_{jtt'}^{+}, q_{jtt'}^{-}, \epsilon, p_{jtt'}$
11: **Repeat** for each product-subperiod ▷ Iterate over j

A.7 Extended Counterfactual Results

A.7.1 Sellers Do Not Update Prices

Table A.7.1: Counterfactuals with No Price Updates, Fixed Marginal Cost

		Recommendation System			
Outcomes	**Baseline**	**Features**	**Query**	**Personalized**	**Most Personalized**
Quantity	508.5	505.9	505.8	505.7	505.8
Gross Booking Revenue ($100s)	1,809.37	1,804.72	1,806.78	1,806.16	1,807.52
Hotel Profits ($100s)	1,019.69	1,019.26	1,019.43	1,018.98	1,019.31
Approx Platform Revenue ($100s)	180.94	180.47	180.68	180.62	180.75
Consumer Welfare					
Δ Consumer Surplus ($100s)	0	86.48	97.18	81.25	40.89
Δ Choice Utility ($100s)	0	93.17	96.47	80.50	33.59
Δ Search Cost ($100s)	0	-6.69	0.71	0.75	7.30

A.7.2 Sellers Update Prices

Table A.7.2: Counterfactuals with Fixed Marginal Costs

Outcomes	Baseline	Recommendation System: Features	Query	Personalized	Most Personalized
Quantity	518.6	496.8	496.2	496.2	495.4
Gross Booking Revenue ($100s)	1,832.93	1,830.82	1,832.84	1,832.78	1,834.03
Hotel Profits ($100s)	1,004.22	1,049.46	1,050.32	1,050.31	1,051.70
Approx Platform Revenue ($100s)	183.29	183.08	183.28	183.28	183.40
Consumer Welfare					
Δ Consumer Surplus ($100s)	0	-37.11	-54.95	-61.11	-87.01
Δ Choice Utility ($100s)	0	-91.53	-116.13	-116.68	-159.24
Δ Search Cost ($100s)	0	54.42	61.18	55.56	72.23

Table A.7.3: Counterfactuals with Common Economies of Scale and Soft-Capacity Constraints

Outcomes	Baseline	Recommendation System: Features	Query	Personalized	Most Personalized
Quantity	519.0	496.2	496.1	495.7	495.3
Gross Booking Revenue ($100s)	1,832.98	1,830.09	1,830.90	1,831.48	1,833.68
Hotel Profits ($100s)	969.88	1,015.12	1,014.21	1,015.42	1,016.96
Approx Platform Revenue ($100s)	183.30	183.01	183.09	183.15	183.37
Consumer Welfare					
Δ Consumer Surplus ($100s)	0	-57.98	-63.06	-74.96	-103.79
Δ Choice Utility ($100s)	0	-116.88	-123.42	-133.56	-175.05
Δ Search Cost ($100s)	0	58.90	60.37	58.60	71.27

APPENDIX B

Appendix to Chapter 2

B.1 Additional Figures

Figure B.1.1: Vacant Appointment Volume RD Estimates by Days in Advance of Appointment

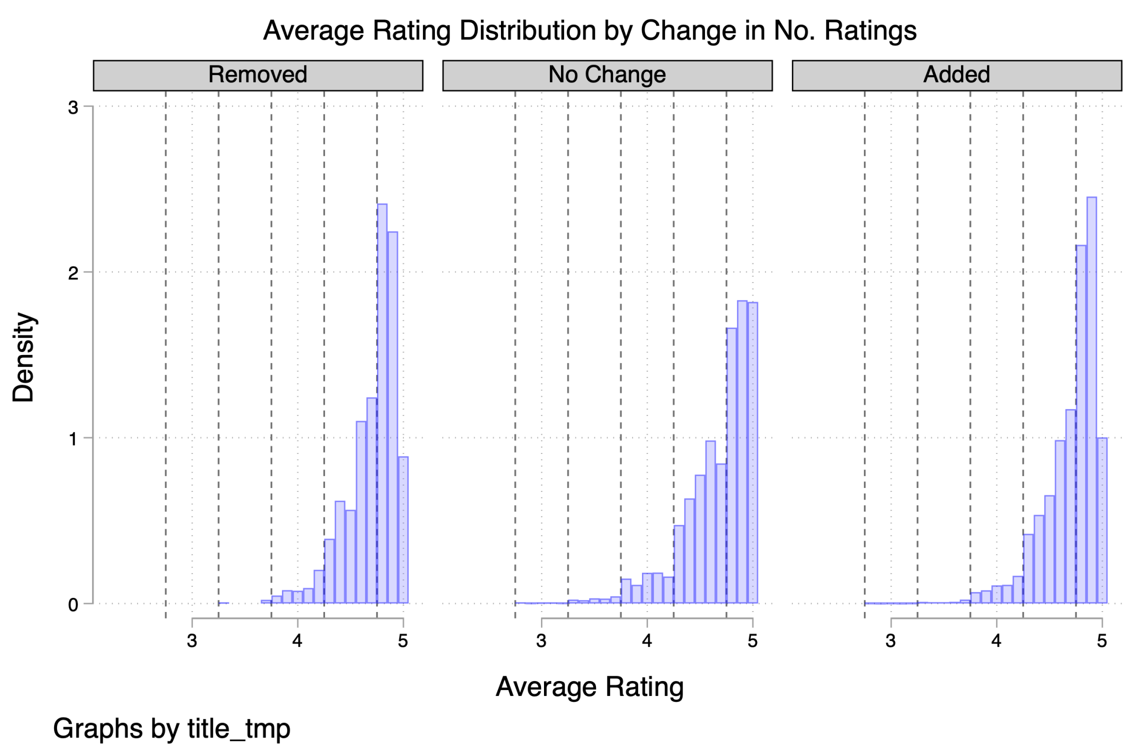


Figure B.1.2: Vacant Appointment Volume RD Estimates by Days in Advance of Appointment

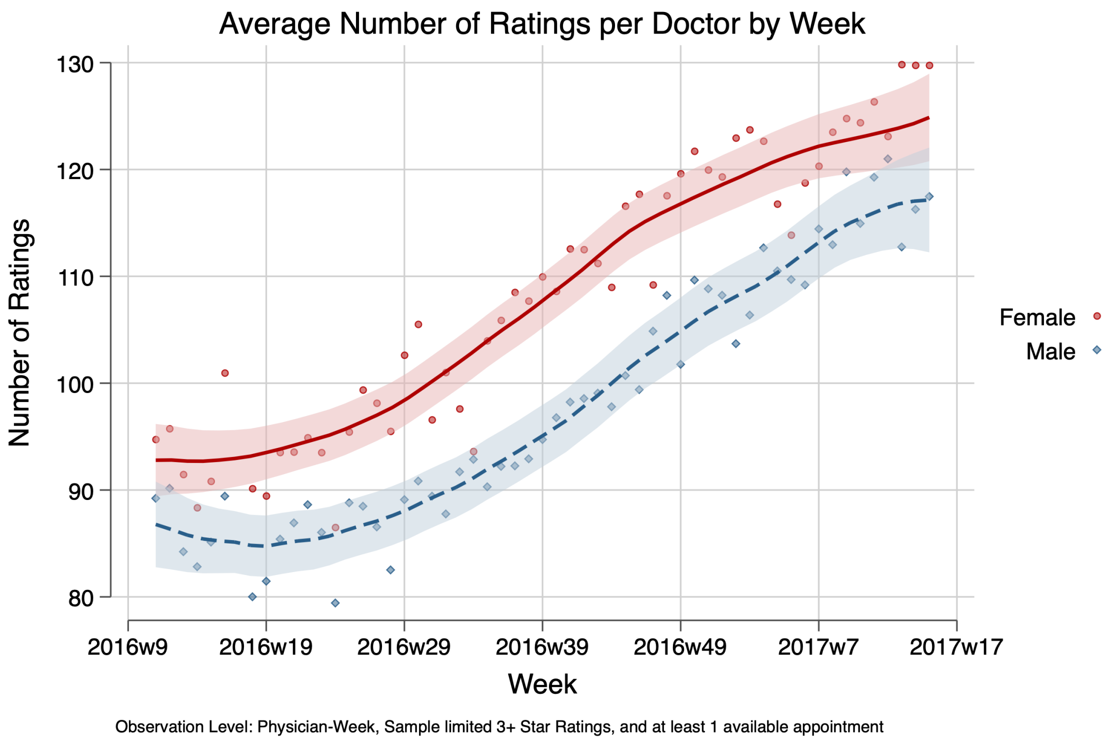


Figure B.1.3: Vacant Appointment Volume RD Estimates by Days in Advance of Appointment

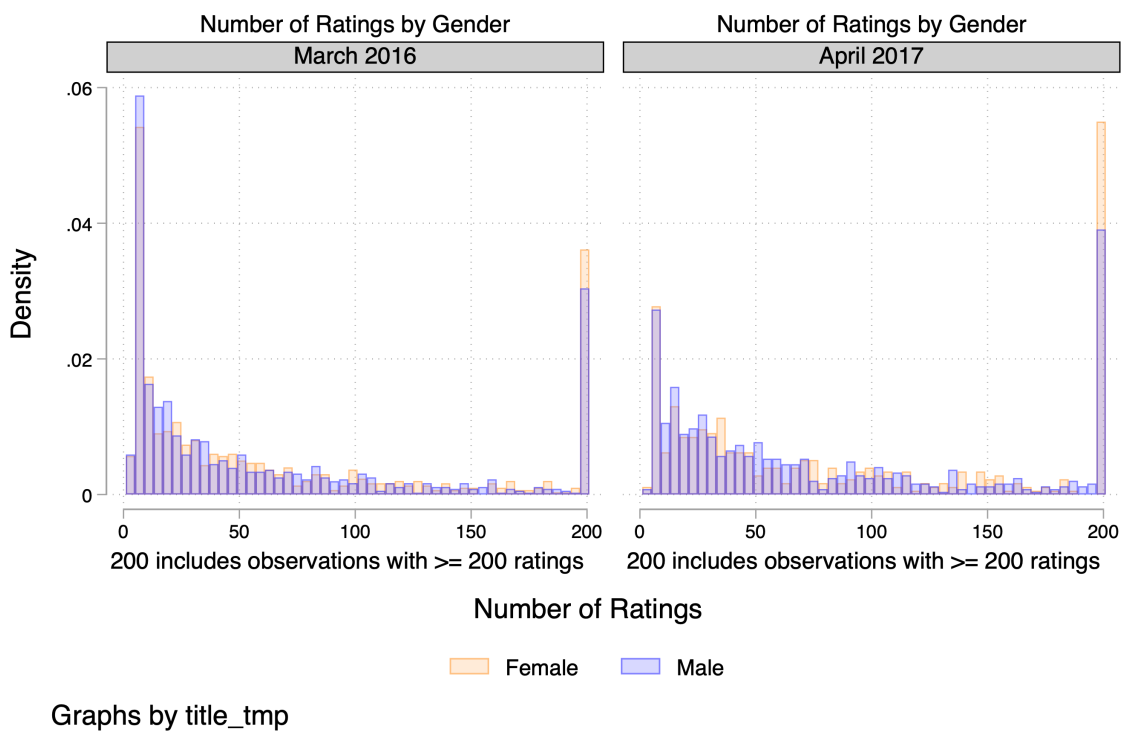
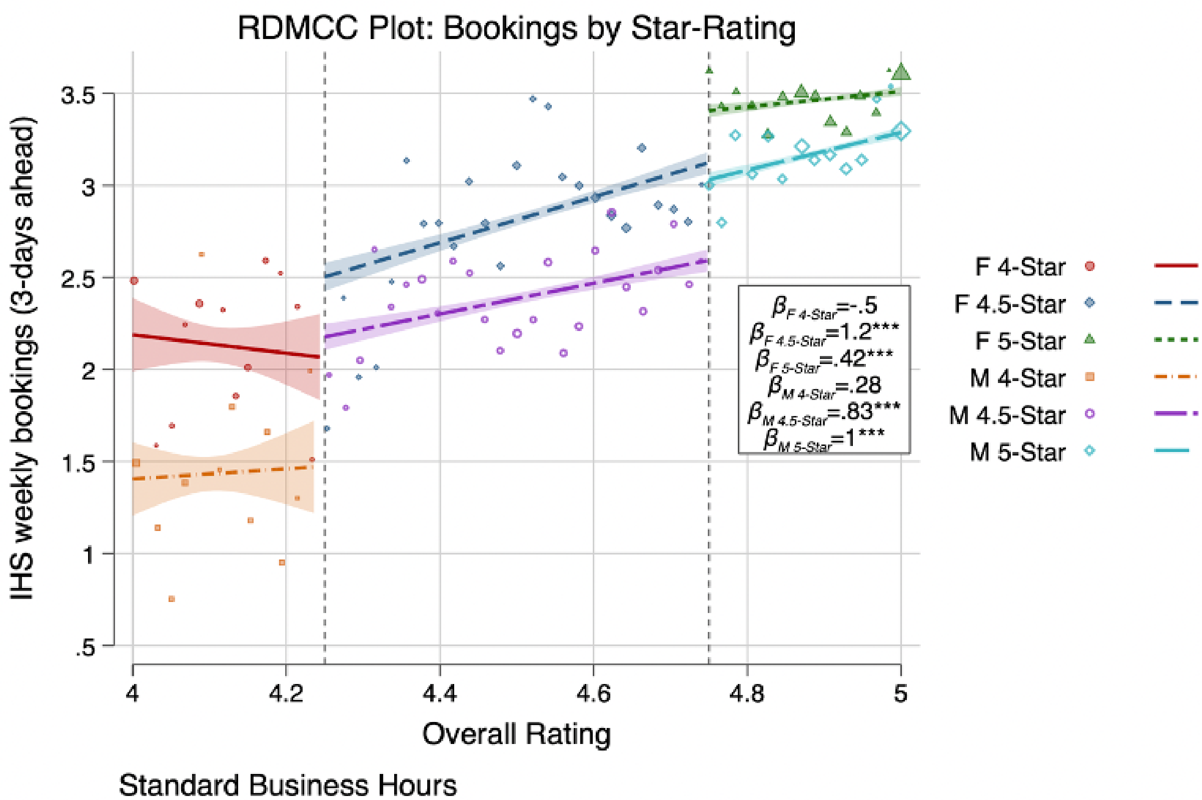


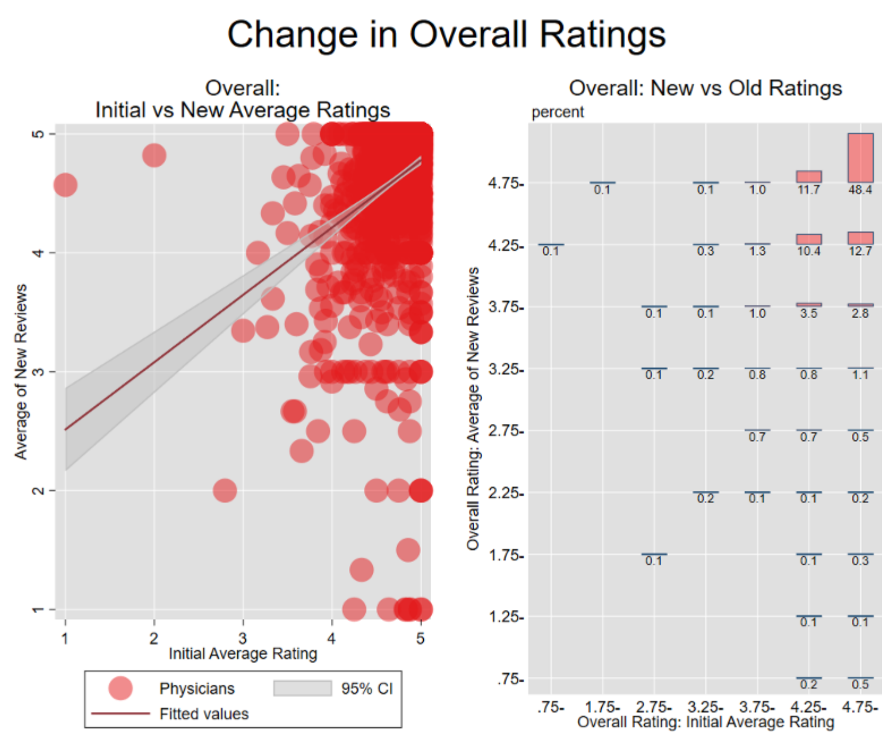
Figure B.1.4: Vacant Appointment Volume RD Estimates by Days in Advance of Appointment



B.2 Changes in Ratings

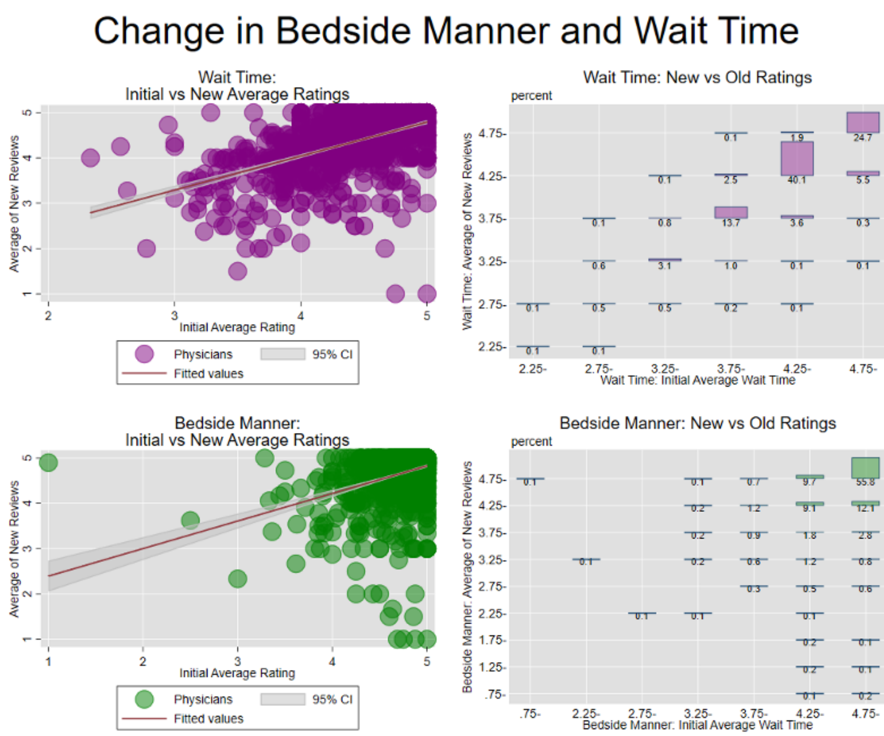
Next, we consider if this platform is causing physicians to improve or change their performance. If patients respond to these ratings, then poorly performing physicians have incentive to change their behavior and the physicians receiving good reviews have incentive to maintain their performance. Additionally, these reviews could be providing feedback the physicians would not otherwise receive. To answer these questions, we look at changes in the ratings over the period of the study. These results are based on a selected a random sample of 1,314 physicians that were persistent in the entire sample. We compute their initial average ratings, which are their overall, bedside manner, and wait time ratings as of February 26, 2016. Next, we compute the averages of all of their new ratings as of February 25, 2017. Figure B.2.1 & Figure B.2.2 illustrate the averages of new ratings (y axis) by initial average rating (x-axis). On the left, there are scatter plots with linear fits of average rating. The right has a two-way tabplot of these same data at the half star level. As an example, on the right plot in Figure B.2.1 , we see that 12.7 percent of physicians had initial overall ratings in the 5-star category and received new ratings that were on average in the 4.5-star category.[1]

Figure B.2.1: Overall Rating Transitions



[1]We round to the nearest half star, so 5-stars would be [4.75, 5], 4.5 star is [4.25, 4.75) and so on.

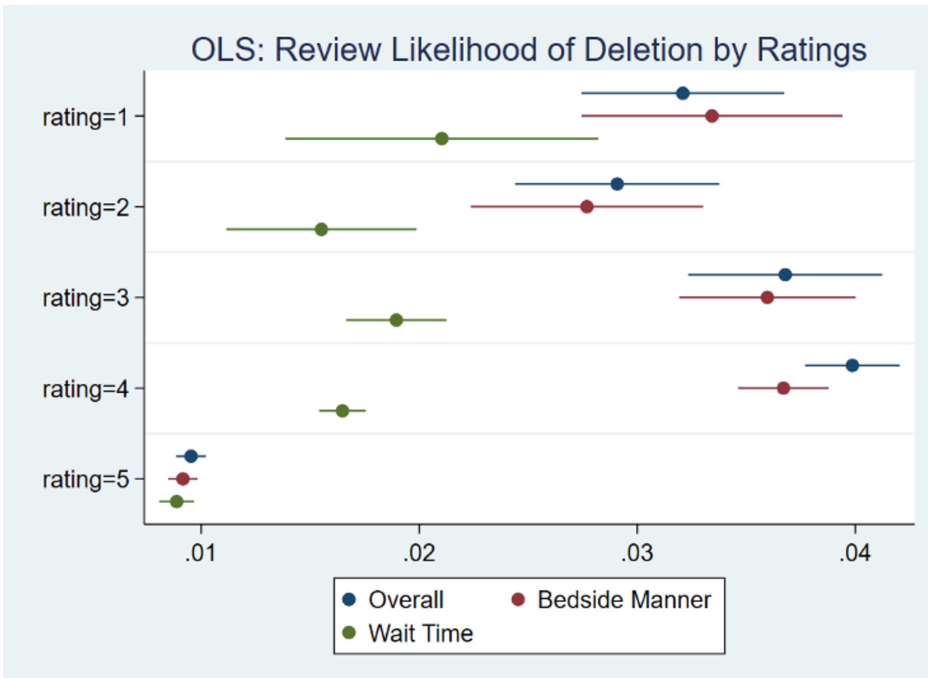
Figure B.2.2: Bedside Manner and Wait Time Rating Transitions



B.3 Deleted Reviews

Here we look at the likelihood of a review being deleted based on the ratings in that review. The results suggest that for overall, bedside manner, and wait time, any review with ratings below 5 stars is more likely to be deleted than 5-star reviews.

Figure B.3.1: Coefficient Plot - Linear Probability of Review Deletion by Rating



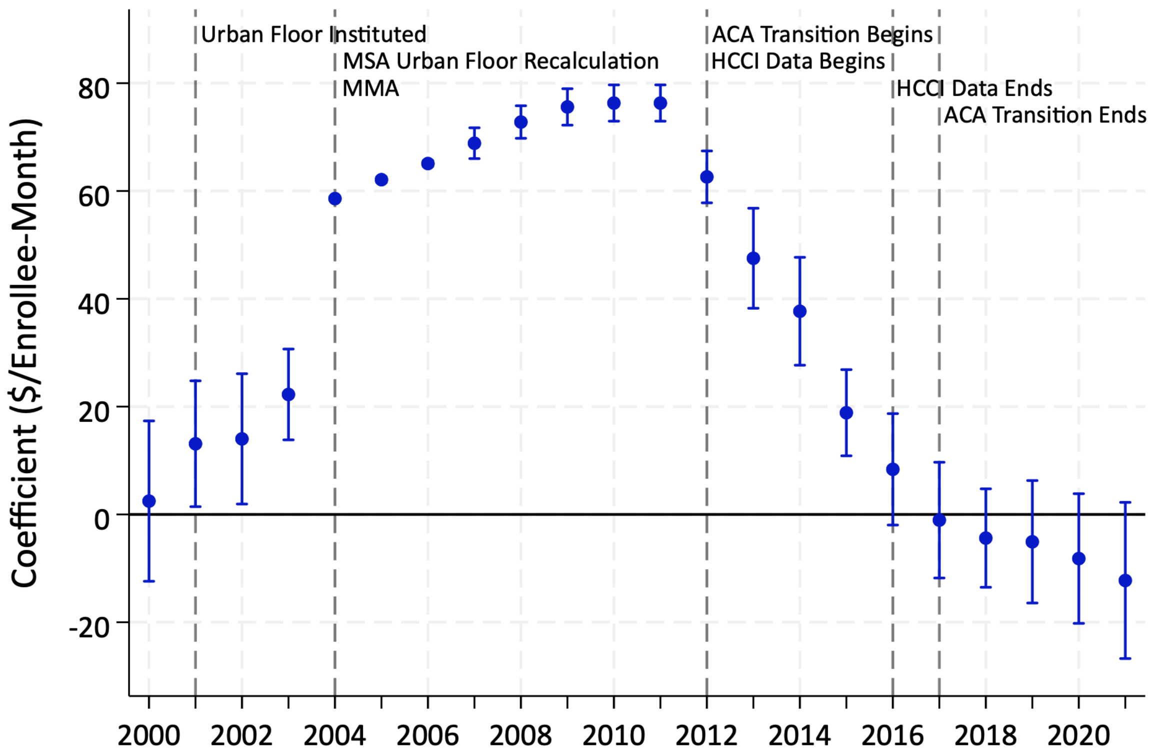
We plan to look into this further, in particular, the interaction terms are interesting. Intuitively one doctor might have an incentive to delete a 4-star review, while another might want to keep it, depending on their average rating, and the bedside manner and wait time ratings in the given reviews.

APPENDIX C

Appendix to Chapter 3

C.1 Additional Tables and Figures

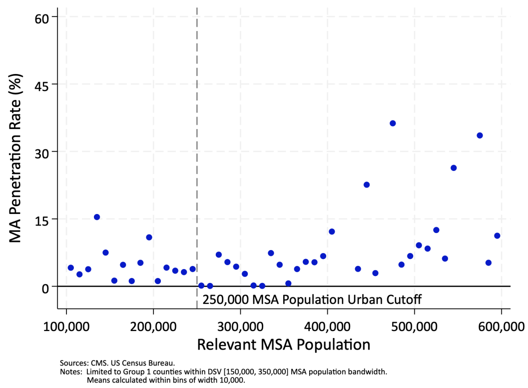
Figure C.1.1: Estimates of Discontinuity in Benchmark Payments by Year without Controls



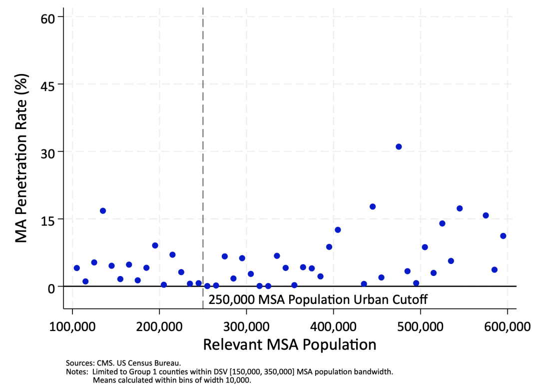
Sources: CMS. US Census Bureau.
Notes: Limited to Group 1 counties within DSV [150,000, 350,000] MSA population bandwidth.
SEs NN clustered at the relevant MSA. Bars represent 95% CIs.

Figure C.1.2: Mean County MA Penetration Rates and Relevant MSA Populations

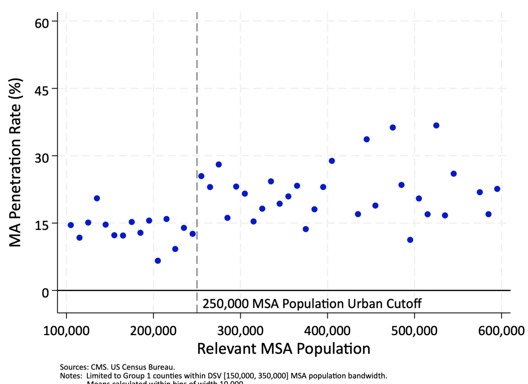
(a) 2000



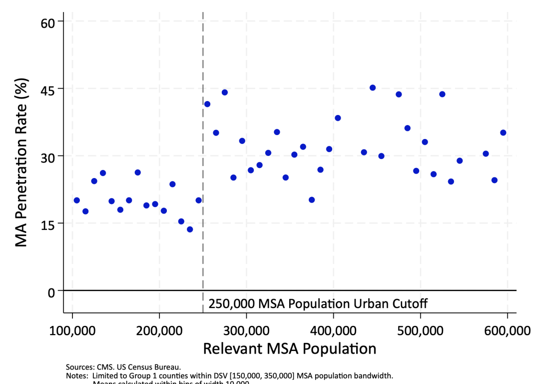
(b) 2003



(c) 2008



(d) 2014



(e) 2019

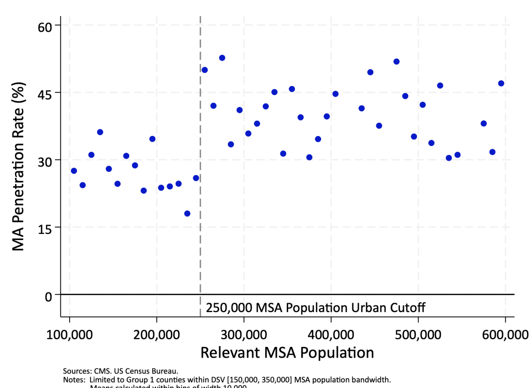
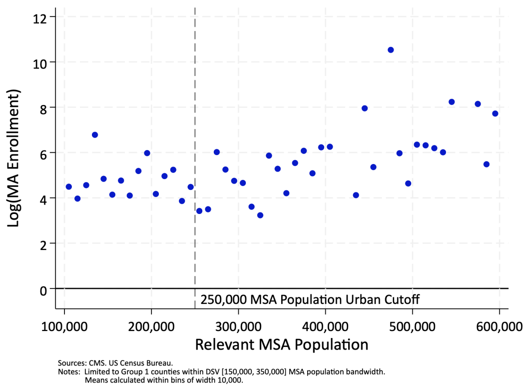
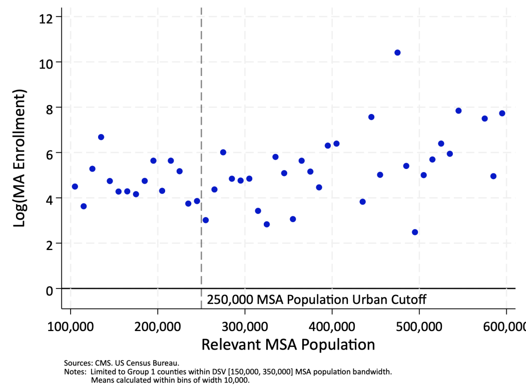


Figure C.1.3: Mean County Log(MA Enrollment) and Relevant MSA Populations

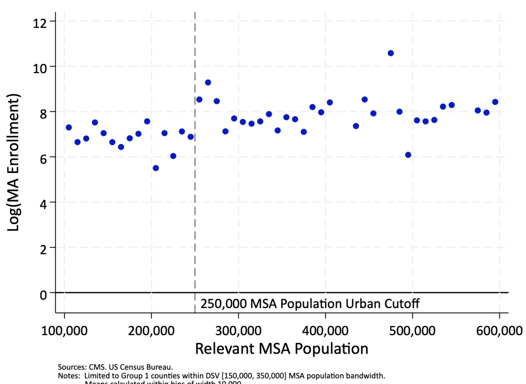
(a) 2000



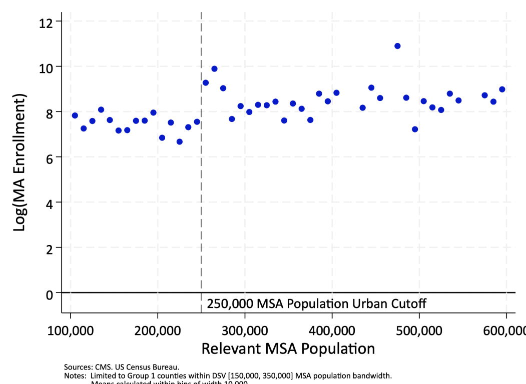
(b) 2003



(c) 2008



(d) 2014



(e) 2019

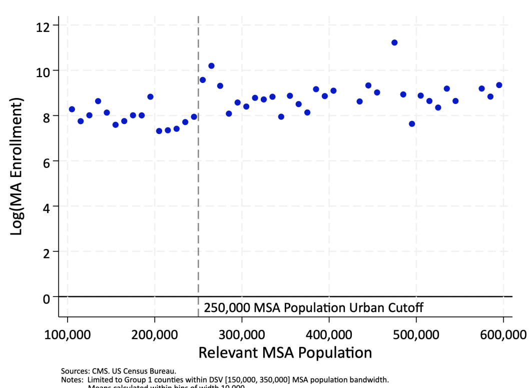
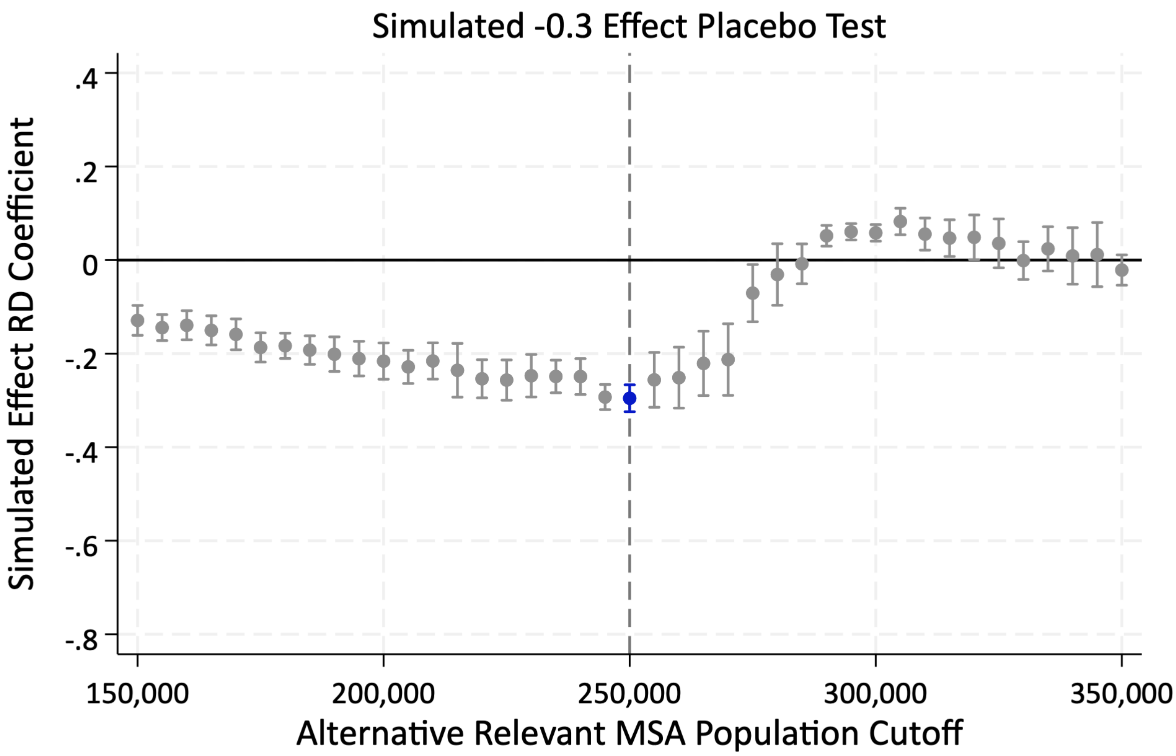


Table C.1.1: RD Estimates of Urban Floor Cutoff on Benchmark Payments, MA Penetration, and Enrollment

MA Benchmark, 2012	63.012	66.407	62.612	66.218
	2.593	2.063	2.454	1.994
Bw	[150000,350000]	[109263.1,3923994]	[150000,350000]	[111392.9,2705669]
N (Left, Right of bw)	[220,180]	[126,369]	[220,180]	[130,356]
MA Penetration Rate, 2012	14.663	14.43	14.54	13.709
	3.122	1.63	3.04	1.624
Bw	[150000,350000]	[109722.7,2752754]	[150000,350000]	[116860.9,2405725]
N (Left, Right of bw)	[220,180]	[126,356]	[220,180]	[139,349]
Log(MA Enrollment), 2012	.869	1.09	1.069	1.232
	.284	.183	.279	.191
Bw	[150000,350000]	[98000.13,2866949]	[150000,350000]	[98810.18,2732342]
N (Left, Right of bw)	[220,180]	[106,363]	[220,180]	[106,356]
MA Benchmark, 2013	48.133	51.502	47.519	50.268
	4.95	3.917	4.734	3.644
Bw	[150000,350000]	[105846,3764248]	[150000,350000]	[107267.1,3072194]
N (Left, Right of bw)	[220,180]	[116,369]	[220,180]	[118,365]
MA Penetration Rate, 2013	15.706	15.089	15.654	14.549
	3.186	1.732	2.966	1.555
Bw	[150000,350000]	[112480.1,2784738]	[150000,350000]	[118187.1,2436197]
N (Left, Right of bw)	[220,180]	[132,356]	[220,180]	[139,349]
Log(MA Enrollment), 2013	.852	1.082	1.063	1.243
	.27	.17	.278	.188
Bw	[150000,350000]	[98197.2,2833470]	[150000,350000]	[99049.01,2763178]
N (Left, Right of bw)	[220,180]	[106,356]	[220,180]	[107,356]
MA Benchmark, 2014	37.582	34.129	37.687	32.565
	5.058	3.668	5.105	3.579
Bw	[150000,350000]	[107675.4,3504013]	[150000,350000]	[114141.2,3198119]
N (Left, Right of bw)	[220,180]	[118,369]	[220,180]	[134,365]
MA Penetration Rate, 2014	15.348	15.543	15.305	15.143
	3.14	2.068	2.947	1.967
Bw	[150000,350000]	[124541.3,2998392]	[150000,350000]	[130002.3,2624697]
N (Left, Right of bw)	[220,180]	[152,365]	[220,180]	[161,355]
Log(MA Enrollment), 2014	.839	1.098	1.051	1.264
	.264	.179	.272	.193
Bw	[150000,350000]	[101223.9,2618965]	[150000,350000]	[100011,2971752]
N (Left, Right of bw)	[220,180]	[109,355]	[220,180]	[108,365]
MA Benchmark, 2015	18.536	12.851	18.862	12.62
	4.016	2.882	4.079	2.679
Bw	[150000,350000]	[88507.84,2821985]	[150000,350000]	[95580.66,2633308]
N (Left, Right of bw)	[220,180]	[83,356]	[220,180]	[96,355]
MA Penetration Rate, 2015	16.418	15.923	16.347	15.361
	3.244	2.343	3.075	2.27
Bw	[150000,350000]	[126799.7,2869409]	[150000,350000]	[131545.9,2551255]
N (Left, Right of bw)	[220,180]	[160,363]	[220,180]	[162,355]
Log(MA Enrollment), 2015	.853	1.106	1.064	1.28
	.265	.183	.277	.2
Bw	[150000,350000]	[99447.59,2679054]	[150000,350000]	[98148.34,2900627]
N (Left, Right of bw)	[220,180]	[107,355]	[220,180]	[106,363]
MA Benchmark, 2016	8.214	9.846	8.368	8.91
	5.081	3.379	5.272	3.225
Bw	[150000,350000]	[89285.48,2697279]	[150000,350000]	[98323.76,2619143]
N (Left, Right of bw)	[220,180]	[92,356]	[220,180]	[106,355]
MA Penetration Rate, 2016	16.571	15.841	16.432	15.312
	3.092	2.18	2.942	2.123
Bw	[150000,350000]	[123653.2,3147875]	[150000,350000]	[129461,2760882]
N (Left, Right of bw)	[220,180]	[152,365]	[220,180]	[161,356]
Log(MA Enrollment), 2016	.781	1.042	.988	1.205
	.247	.155	.265	.184
Bw	[150000,350000]	[96975.75,2622403]	[150000,350000]	[96817.98,2972065]
N (Left, Right of bw)	[220,180]	[103,355]	[220,180]	[103,365]
DSV Bandwidth	X	X		
Data Driven BW			X	X
Controls	X		X	

Sources: CMS. US Census Bureau.
Notes: Limited to Group 1 counties.
SEs NN clustered at the relevant MSA.
RD Controls: Log(1997 FFS Rate), Log(1997 MA Enrollment).

Figure C.1.4: Placebo Test of a Simulated -0.3 Effect at Cutoff



Sources: Authors' calculations from HCCI. CMS. US Census Bureau.
Notes: RD estimates and SEs result from estimating RD using Group 1 counties and data-driven bandwidths at selected cutoff.
SEs NN clustered at the relevant MSA. Bars represent 95% CIs.
Data simulated. Counties w/ MSA pop < 250k value = N(0, 0.1). Counties w/ MSA pop > 250k = -0.3 + N(0, 0.1).

Bibliography

Abadie, A. and G. W. Imbens (2008). Estimation of the conditional variance in paired experiments. *Annales d'Economie et de Statistique 91/92*, 175–187.

Abaluck, J., G. Compiani, and F. Zhang (2020, March). A Method to Estimate Discrete Choice Models that is Robust to Consumer Search. Technical Report w26849, National Bureau of Economic Research, Cambridge, MA.

Abaluck, J. and J. Gruber (2016). Evolving choice inconsistencies in choice of prescription drug insurance. *American Economic Review 106*(8), 2145–2184.

Afendulis, C. C., M. E. Chernew, and D. P. Kessler (2017). The effect of medicare advantage on hospital admissions and mortality. *American Journal of Health Economics 3*(2), 254–279.

Agrawal, K., S. Athey, A. Kanodia, and E. Palikot (2022, December). Personalized Recommendations in EdTech: Evidence from a Randomized Controlled Trial. arXiv:2208.13940 [econ, q-fin].

Akerlof, G. A. (1970, August). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics 84*(3), 488.

Andrews, I., M. Gentzkow, and J. M. Shapiro (2020, October). Transparency in Structural Research. *Journal of Business & Economic Statistics 38*(4), 711–722.

Athey, S. and G. Ellison (2011). Position Auctions with Consumer Search. *The Quarterly Journal of Economics 126*(3), 1213–1270. Publisher: Oxford University Press.

Baker, L. C., M. K. Bundorf, A. M. Devlin, and D. P. Kessler (2016). Medicare advantage plans pay hospitals less than traditional medicare pays. *Health Affairs 35*(8), 1444–1451.

Baker, L. C., M. K. Bundorf, and D. P. Kessler (2020). The effects of medicare advantage on opioid use. *Journal of Health Economics 70*, 102278.

Berry, S., J. Levinsohn, and A. Pakes (1995, July). Automobile Prices in Market Equilibrium. *Econometrica 63*(4), 841.

Berry, S. T. (1994). Estimating Discrete-Choice Models of Product Differentiation. *The RAND Journal of Economics 25*(2), 242.

Betancourt, J., A. Hortaçsu, A. Oery, and K. Williams (2022, August). Dynamic Price Competition: Theory and Evidence from Airline Markets. Technical Report w30347, National Bureau of Economic Research, Cambridge, MA.

Blake, T., S. Moshary, K. Sweeney, and S. Tadelis (2021, July). Price Salience and Product Choice. *Marketing Science 40*(4), 619–636.

Brown, Z., C. Hansman, J. Keener, and A. Veiga (2023, March). Information and Disparities in Health Care Quality: Evidence from GP Choice in England. Technical Report w31033, National Bureau of Economic Research, Cambridge, MA.

Brown, Z. Y. and J. Jeon (2022). Endogenous Information and Simplifying Insurance Choice.

Buchmueller, T., A. Kaye, W. Mandelkorn, and S. Miller (2022). How do medicare advantage prices vary geographically? evidence from a new price index.

Bundorf, M. K., N. Chun, G. S. Goda, and D. P. Kessler (2009, May). Do markets respond to quality information? The case of fertility clinics. *Journal of Health Economics 28*(3), 718–727.

Burges, C. J. C. (2010). From RankNet to LambdaRank to LambdaMART: An Overview. *Learning 11.*

Cabral, M., M. Geruso, and N. Mahoney (2018, August). Do larger health insurance subsidies benefit patients or producers? evidence from medicare advantage. *American Economic Review 108*(8), 2048–87.

Cardell, N. S. (1997, April). Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity. *Econometric Theory 13*(2), 185–213.

Cattaneo, M. D., R. Titiunik, and G. Vazquez-Bare (2020, December). Analysis of regression-discontinuity designs with multiple cutoffs or multiple scores. *The Stata Journal: Promoting communications on statistics and Stata 20*(4), 866–891.

Chan, A. (2023). Discrimination Against Doctors: A Field Experiment.

Chapelle, O. and Y. Chang (2011, January). Yahoo! Learning to Rank Challenge Overview. In *Proceedings of the Learning to Rank Challenge*, pp. 1–24. PMLR. ISSN: 1938-7228.

Chen, Y. and S. Yao (2017, December). Sequential Search with Refinement: Model and Application with Click-Stream Data. *Management Science 63*(12), 4345–4365.

Choi, M., A. Y. Dai, and K. Kim (2018). Consumer Search and Price Competition. *Econometrica 86*(4), 1257–1281.

Compiani, G., G. Lewis, S. Peng, and W. Wang (2021). Online Search and Product Rankings: A Double Index Approach. *SSRN Electronic Journal*.

Cooper, Z., S. V. Craig, M. Gaynor, and J. Van Reenen (2018, 09). The Price Ain't Right? Hospital Prices and Health Spending on the Privately Insured*. *The Quarterly Journal of Economics 134*(1), 51–107.

Curto, V., L. Einav, A. Finkelstein, J. Levin, and J. Bhattacharya (2019, April). Health care spending and utilization in public and private medicare. *American Economic Journal: Applied Economics 11*(2), 302–32.

Cutler, D. M., R. S. Huckman, and M. B. Landrum (2004, April). The Role of Information in Medical Markets: An Analysis of Publicly Reported Outcomes in Cardiac Surgery. *American Economic Review 94*(2), 342–346.

Dafny, L. and D. Dranove (2008, September). Do report cards tell consumers anything they don't already know? The case of Medicare HMOs. *The RAND Journal of Economics 39*(3), 790–821.

Dafny, L., K. Ho, and E. Kong (2022). How do copayment coupons affect branded drug prices and quantities purchased? Working Paper 29735, National Bureau of Economic Research.

Darby, M. R. and E. Karni (1973). Free Competition and the Optimal Amount of Fraud. *The Journal of Law & Economics 16*(1), 67–88. Publisher: [University of Chicago Press, Booth School of Business, University of Chicago, University of Chicago Law School].

De los Santos, B. and S. Koulayev (2016, July). Optimizing Click-Through in Online Rankings with Endogenous Search Refinement.

Diamond, P. A. (1971, June). A model of price adjustment. *Journal of Economic Theory 3*(2), 156–168.

Dinerstein, M., L. Einav, J. Levin, and N. Sundaresan (2018, July). Consumer Price Search and Platform Design in Internet Commerce. *American Economic Review 108*(7), 1820–1859.

Donnelly, R., A. Kanodia, and I. Morozov (2020). Welfare Effects of Personalized Rankings. *SSRN Electronic Journal*.

Dranove, D., D. Kessler, M. McClellan, and M. Satterthwaite (2003, June). Is More Information Better? The Effects of "Report Cards" on Health Care Providers. *Journal of Political Economy 111*(3), 555–588.

Duggan, M., A. Starc, and B. Vabson (2016). Who benefits when the government pays more? pass-through in the medicare advantage program. *Journal of Public Economics 141*, 50–67.

Ellison, G. and S. F. Ellison (2009). Search, Obfuscation, and Price Elasticities on the Internet. *Econometrica 77*(2), 427–452.

Farronato, C. and A. Fradkin (2022, June). The Welfare Effects of Peer Entry: The Case of Airbnb and the Accommodation Industry. *American Economic Review 112*(6), 1782–1817.

Farronato, C., A. Fradkin, and A. MacKay (2023, January). Self-Preferencing at Amazon: Evidence from Search Rankings. Technical Report w30894, National Bureau of Economic Research, Cambridge, MA.

Feng Lu, S. (2012). Multitasking, Information Disclosure, and Product Quality: Evidence from Nursing Homes. *Journal of Economics & Management Strategy 21*(3), 673–705. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1530-9134.2012.00341.x.

Fink, M., K. Klein, K. Sayers, J. Valentino, C. Leonardi, A. Bronstone, P. M. Wiseman, and V. Dasa (2020, January). Objective Data Reveals Gender Preferences for Patients' Primary Care Physician. *Journal of Primary Care & Community Health 11*, 2150132720967221. Publisher: SAGE Publications Inc.

Galichon, A. (2022, April). ON THE REPRESENTATION OF THE NESTED LOGIT MODEL. *Econometric Theory 38*(2), 370–380.

Gardete, P. M. and M. Hunter (2018). Avoiding Lemons in Search of Peaches: Designing Information Provision. Research Papers, Stanford University, Graduate School of Business.

Gaynor, M., K. Ho, and R. J. Town (2015, June). The industrial organization of health-care markets. *Journal of Economic Literature 53*(2), 235–84.

Geddes, E. (2022, December). The effects of price regulation in markets with strategic entry: Evidence from health insurance markets. Working paper.

Goldfarb, A. and C. Tucker (2019, March). Digital Economics. *Journal of Economic Literature 57*(1), 3–43.

Gowrisankaran, G., A. Nevo, and R. Town (2015, January). Mergers when prices are negotiated: Evidence from the hospital industry. *American Economic Review 105*(1), 172–203.

Greminger, R. P. (2022). Heterogeneous Position Effects and the Power of Rankings. Publisher: arXiv Version Number: 2.

Handel, B. and J. Kolstad (2015). Getting the most from marketplaces: Smart policies on health insurance choice. *Brookings Hamilton Project: Discussion Paper 2015-08*.

Handel, B. R. (2013, December). Adverse selection and inertia in health insurance markets: When nudging hurts. *American Economic Review 103*(7), 2643–82.

Honka, E. (2014, December). Quantifying search and switching costs in the US auto insurance industry. *The RAND Journal of Economics 45*(4), 847–884.

Honka, E. and P. Chintagunta (2017, January). Simultaneous or Sequential? Search Strategies in the U.S. Auto Insurance Industry. *Marketing Science 36*(1), 21–42.

Hortaçsu, A., O. Natan, H. Parsley, T. Schwieg, and K. Williams (2021, December). Demand Estimation with Infrequent Purchases and Small Market Sizes. Technical Report w29530, National Bureau of Economic Research, Cambridge, MA.

Hu, H. and J. Pang (2021). Model Extraction and Defenses on Generative Adversarial Networks. Publisher: arXiv Version Number: 1.

Jin, G. Z. and A. T. Sorensen (2006, March). Information and consumer choice: The value of publicized health plan ratings. *Journal of Health Economics 25*(2), 248–275.

Kim, J. B., P. Albuquerque, and B. J. Bronnenberg (2010). Online Demand Under Limited Consumer Search. *Marketing Science 29*(6), 1001–1023. Publisher: INFORMS.

Koulayev, S. (2014). Search for differentiated products: identification and estimation. *The RAND Journal of Economics 45*(3), 553–575. Publisher: Wiley.

Lam, H. T. (2021). Platform Search Design and Market Power.

League, R. J., P. Eliason, R. C. McDevitt, J. W. Roberts, and H. Wong (2022, 02). Variability in Prices Paid for Hemodialysis by Employer-Sponsored Insurance in the US From 2012 to 2019. *JAMA Network Open 5*(2), e220562–e220562.

Lee, K. H. and L. Musolff (2021). Entry Into Two-Sided Markets Shaped By Platform-Guided Search.

Lin, E., B. Ly, E. Duffy, and E. Trish (2022). Medicare advantage plans pay large markups to consolidated dialysis organizations. *Health Affairs 41*(8), 1107–1116.

Luca, M. (2011). Reviews, Reputation, and Revenue: The Case of Yelp.Com. *SSRN Electronic Journal*.

Maeda, J. and L. N. L (2018, Dec). How do the hospital prices paid by medicare advantage plans and commercial plans compare with medicare fee-for-service prices? *Inquiry 55*.

Marzilli Ericson, K. M. (2014, February). Consumer inertia and firm pricing in the medicare part d prescription drug insurance exchange. *American Economic Journal: Economic Policy 6*(1), 38–64.

McCarthy, I. and M. V. Raval (2022, September). Price spillovers and specialization in health care: The case of children's hospitals. Working Paper 30425, National Bureau of Economic Research.

McCarthy, I., K. Sanbower, and L. S. Aragon (2022). Online Reviews and Hospital Choices.

McCrary, J. (2008, February). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics 142*(2), 698–714.

McFadden, D. (1989). A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration. *Econometrica 57*(5), 995–1026. Publisher: [Wiley, Econometric Society].

Meyers, D. J., V. Mor, and M. Rahman (2018). Medicare advantage enrollees more likely to enter lower-quality nursing homes compared to fee-for-service enrollees. *Health Affairs 37*(1), 78–85.

Moehring, A. (2023). Personalized Rankings and User Engagement: An Empirical Evaluation of the Reddit News Feed.

Morozov, I. (2023, March). Measuring Benefits from New Products in Markets with Information Frictions. *Management Science*, mnsc.2023.4729.

Morozov, I., S. Seiler, X. Dong, and L. Hou (2021, September). Estimation of Preference Heterogeneity in Markets with Costly Search. *Marketing Science 40*(5), 871–899.

Norton, E. C. (2022, September). The inverse hyperbolic sine transformation and retransformed marginal effects. *The Stata Journal: Promoting communications on statistics and Stata 22*(3), 702–712.

Nosal, K. (2011, October). Estimating switching costs for medicare advantage plans. Working paper.

Oh, S. J., M. Augustin, B. Schiele, and M. Fritz (2018, February). Towards Reverse-Engineering Black-Box Neural Networks. arXiv:1711.01768 [cs, stat].

Orekondy, T., S. J. Oh, Y. Zhang, B. Schiele, and M. Fritz (2020, September). Gradient-Leaks: Understanding and Controlling Deanonymization in Federated Learning. arXiv:1805.05838 [cs, stat].

Papernot, N., P. McDaniel, I. Goodfellow, S. Jha, Z. B. Celik, and A. Swami (2017, March). Practical Black-Box Attacks against Machine Learning. arXiv:1602.02697 [cs].

Platt, J. (2000). Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. *Adv. Large Margin Classif. 10*.

Reimers, I. and J. Waldfogel (2023, October). A Framework for Detection, Measurement, and Welfare Analysis of Platform Bias. Technical Report w31766, National Bureau of Economic Research, Cambridge, MA.

Revelt, D. and K. Train (2000). Customer-Specific Taste Parameters and Mixed Logit.

Rhodes, J. (2020). Private health insurer incentives and prescription opioid use: Evidence from medicare part d. Dissertation, University of Michigan; https://deepblue.lib.umich.edu/bitstream/handle/2027.42/163252/jhrhodes$_1.pdf?sequence = 1isAllowed = y$.

Sarsons, H. Interpreting Signals in the Labor Market: Evidence from Medical Referrals.

Song, Z., M. B. Landrum, and M. E. Chernew (2013). Competitive bidding in medicare advantage: Effect of benchmark changes on plan bids. *Journal of Health Economics 32*(6), 1301–1312.

Stigler, G. J. (1961, June). The Economics of Information. *Journal of Political Economy 69*(3), 213–225.

Strombom, B. A., T. C. Buchmueller, and P. J. Feldstein (2002). Switching costs, price sensitivity and health plan choice. *Journal of Health Economics 21*(1), 89–116.

Teng, X. (2022). Self-Preferencing, Quality Provision, and Welfare in Mobile Application Markets. *SSRN Electronic Journal*.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

Tramer, F., F. Zhang, A. Juels, M. K. Reiter, and T. Ristenpart (2016). Stealing Machine Learning Models via Prediction APIs. *25th USENIX Security Symposium (USENIX Security 16)*.

Ursu, R., S. Seiler, and E. Honka (2023, January). The Sequential Search Model: A Framework for Empirical Research.

Ursu, R. M. (2018, August). The Power of Rankings: Quantifying the Effect of Rankings on Online Consumer Search and Purchase Decisions. *Marketing Science 37*(4), 530–552.

Wang, J., J. Hockenberry, S.-Y. Chou, and M. Yang (2011, March). Do bad report cards have consequences? Impacts of publicly reported provider quality information on the CABG market in Pennsylvania. *Journal of Health Economics 30*(2), 392–407.

Weitzman, M. L. (1979, May). Optimal Search for the Best Alternative. *Econometrica 47*(3), 641.

Werner, R. M., R. T. Konetzka, E. A. Stuart, E. C. Norton, D. Polsky, and J. Park (2009, August). Impact of Public Reporting on Quality of Postacute Care. *Health Services Research 44*(4), 1169–1187.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed ed.). Cambridge, Mass: MIT Press. OCLC: ocn627701062.

Xu, J. and D. Polsky (2023). Comparing medicare advantage and traditional medicare prices for hospital outpatient services with hospital price transparency data. *Medical Care Res Rev. 80*(4), 455–461.