# Four Essays on the Economics of Digitization

vorgelegt von
M.Sc.
Maximilian Schäfer

an der Fakultät VII - Wirtschaft und Management
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Wirtschaftswissenschaften
- Dr. rer. oec. -

genehmigte Dissertation

Promotionsausschuss:

| | |
|---|---|
| Vorsitzende: | Prof. Dr. Radosveta Ivanova-Stenzel |
| Gutachter: | Prof. Dr. Tomaso Duso |
| | Prof. Dr. Hannes Ullrich |

Tag der wissenschaftlichen Aussprache: 24.11.2020

Berlin, 2020

# Abstract

Technological progress in information technologies brings unprecedented economic opportunities. By drastically reducing costs, it has never been easier to access information, communicate, and to exchange goods and services. At the same time, the ensuing rapid change of the economic landscape raises numerous challenges that deserve special scrutiny from the economic profession. The present dissertation addresses some economic implications of the digital transformation in four separate articles. The first chapter analyzes the role of data in the success of firms in digital markets. Extracting value from data by relying on algorithms with an unprecedented ability to predict individuals' preferences and choices is the core business model of many digital firms. Yet, from an economic perspective, the value of data is controversially debated. We present empirical evidence that data do not constitute a conventional input, highlighting a mechanism that sets data apart from more conventional sources of market power. The second article deals with the disruptive impact of digital peer-to-peer platforms for consumers and the resulting externalities for established industries. Using a discrete choice model of demand, we quantify the welfare impact of Airbnb's entry in the short-term accommodation market. The third article studies a further externality caused by peer-to-peer markets. Using plausibly exogeneous variation, it quantifies the impact of Airbnb on rents in Berlin, Germany. The last article analyzes the role of reviews and ratings in digital marketplaces. Using transaction and rating data on Airbnb, it provides empirical support for the hypothesis that ratings are influenced by prices and that sellers use this mechanism to generate additional revenues.

# Zusammenfassung

Der technische Fortschritt in den Informationstechnologien eröffnet beispiellose wirtschaftliche Möglichkeiten. Durch eine drastische Reduzierung der Kosten dieser Technologien ist es heute so einfach wie noch nie, Informationen und Wissen zu akquirieren, zu kommunizieren und Produkte und Dienstleistungen zu handeln. Die damit einhergehende Umwälzung der Wirtschaft wirft zahlreiche ökonomische Fragestellungen auf. Die vorliegende Dissertation umfasst vier Abschnitte, welche einen Teil dieser Fragestellungen beleuchten. Der erste Abschnitt analysiert, inwiefern Daten eine marktbeherrschende Stellung begünstigen. Die Haupteinkommensquelle zahlreicher digitaler Unternehmen beruht auf der massenhaften automatisierten Auswertung personenbezogener Daten mit Hilfe komplexer Algorithmen. Nichtsdestotrotz ist der wirtschaftliche Wert von Daten, insbesondere im Vergleich zu konventionellen Produktionsfaktoren, unter Ökonomen sehr umstritten. Wir legen empirische Evidenz vor, dass Daten eine marktbeherrschende Stellung durch eine neue Externalität begünstigen, die bei konventionellen Produktionsfaktoren nicht vorliegt. Der zweite Abschnitt befasst sich mit den disruptiven Auswirkungen der Plattformökonomie auf Verbraucher und den damit einhergehenden Externalitäten für etablierte Industrien. Mithilfe eines strukturellen empirischen Modells werden die Wohlfahrtswirkungen Airbnbs auf den Beherbergungssektor quantifiziert. Der dritte Abschnitt analysiert eine weitere Externalität, die durch die Plattformökonomie verursacht wird. Mit Hilfe plausibler exogener Variationen werden die Auswirkungen von Airbnb auf die Langzeitmieten ermittelt. Der letzte Abschnitt untersucht die Rolle von Kundenbewertungen in Online-Märkten und legt empirische Evidenz vor, dass Gastgeber durch Preissetzung Bewertungen positiv beeinflussen und hierdurch ihr Einkommen steigern können.

# Acknowledgements

Completing this dissertation would not have been possible alone. In the years since I started my journey, I have lived through the ups and downs of academic research. The experience has instilled a tremendous amount of humility and humbleness in me toward the process of scientific discovery. I am grateful to everyone who patiently accompanied me along this way.

I am especially grateful to Tomaso Duso, who believed in me from the beginning and who accompanied every step of this dissertation. I am thankful that I had the luck to become his doctoral student. He opened many doors - I only had to walk through. I am also extremely grateful to Hannes Ullrich, who dedicated a lot of his time, knowledge, and expertise to my scientific endeavor. His friendly advice contributed enormously to the quality of this dissertation. I am also very lucky that I could rely on the knowledge, expertise, and friendship of Kevin Tran with whom I worked together during these past years.

I am immensely thankful to my parents who raised me to become a curious, open-minded, and independent individual. Without their love and support, none of this would have been possible. I am also immensely grateful and lucky that I could always count on the unconditional and loving support of Vera who faithfully stands by my side in every situation. Without Marica, Martin, and Patrick life during the PhD would not have been half as colorful – I thank them for their many wonderful moments of companionship. A very special thanks goes to every PhD student of the 2015 cohort at the DIW Graduate Center and to my colleagues at the Firms and Markets Department: You made my PhD years a memorable experience.

I am thankful to DIW Berlin, the Berlin Center for Consumer Policy, and the Fritz-Thyssen Stiftung for their financial support. A special thanks goes to every member of DIW Graduate Center and the Firms and Market Department, particularly Irene Faas, Juliane Metzner, Sibylle Kremser, and Sandra Schreiber for their help in administrative duties. I thank Adam Lederer for proofreading. I thank Claus Michelsen for offering me the opportunity to gain policy experience. I thank Martin Gornig for unbureaucratic help. Last, but not least, I also thank Geza Sapi for always carrying on with our research project.

# Rechtliche Erklärung

Hiermit versichere ich, Maximilian Schäfer, dass ich die vorliegende Arbeit selbstständig und ohne unzulässige Hilfsmittel verfasst habe. Die verwendeten Quellen sind vollständig im Literaturverzeichnis angegeben. Die Arbeit wurde noch keiner Prüfungsbehörde in gleicher oder ähnlicher Form vorgelegt.

Maximilian Schäfer

Bologna, den 10. Dezember 2020

# Table of Contents

# List of Tables

# List of Figures

# List of Abrreviations

**2SLS** Two-Stage Least Squares

**AHLA** American Hotel & Lodging Association

**DCG** Discounted Cumulative Gain

**e.g.** exempli gratia (for example)

**FE** fixed effects

**i.e.** id est (that is)

**INSEE** National Institute of Statistics and Economic Studies

**IV** instrumental variable

**LOR** Lebensweltlich orientierte Räume

**max.** maximum

**min.** minimum

**MS** market size

**NL** nested logit

**OR** occupancy ratio

**std.** standard deviation

**URL** Uniform Resource Locator

**vs.** versus

**ZwVbG** Zweckentfremdungsverbot-Gesetz

# Chapter 1

# Introduction

## 1.1 Economic Challenges of the Digital Transformation

The ascent of digital technologies caused by the massive reduction in the costs of computing, data-storage, and communication is at the core of much of the socioeconomic transformations that we currently witness. The widespread availability of computing devices and new means of communication has fundamentally changed the way people exchange information, goods, and services. The increased power of the companies driving these changes and the externalities they impose on other industries and society in general often pose fundamental challenges to policy makers. This dissertation addresses some economic policy challenges related to the digital transformation of the economy in four separate chapters.

The second chapter of the dissertation analyzes the implications of the increasing importance of data for competition policy. The rise of superstar firms in digital industries has put data on the center stage of the policy debate (The Economist, 2017, 2018). Services like Google, Twitter, Facebook, Instagram, and TikTok derive a large share of their revenues from the monetization of their user data, which is mainly achieved through advertising. Learning about user preferences through data is key to maximize revenues. Additionally, improvement of service quality through learning from data is essential to keep users engaged and to acquire new users.

Naturally, the question arises to what extent the ownership of large data troves gives a firm a competitive advantage over potential rivals. Is it realistic to assume that an entrant with an in principle superior technology is able to unsettle an established incumbent without having access to any data to begin with? While the continually increasing market power of firms in data-driven industries suggests data shield firms from competition, the question remains controversially debated because of the lack of a clear causal mechanism (Newman, 2014; Grunes and Stucke, 2015; Schepp and Wambach, 2015; Tucker, 2019). The second chapter of this dissertation analyzes returns form data using search engine data from Yahoo! and highlights a mechanism that clearly supports the hypothesis that data confer a competitive advantage.

The third and fourth chapters of the dissertation address the issue of economic externalities caused by the digitization of the economy (Einav *et al.*, 2016). The rise of digital platforms has opened new possibilities for individuals to exchange goods and services though online marketplaces (Sundararajan, 2017). Because sellers and buyers on these platforms usually consist of simple individuals, they are often characterized as peer-to-peer platforms. Peer-to-peer platforms are often viewed as beneficial for society because they allow to productively use idle assets like spare goods (eBay), space (Airbnb) or time (Uber, TaskRabbit).

Due to their widespread adoption, mostly unhindered by regulatory constraints, some peer-to-peer platforms have quickly grown into major players (Vaughan and Daverio, 2016). Their rise has made apparent their sometimes far reaching consequences for established industries and society (Drahokoupil and Fabo, 2016; Sundararajan, 2017; Geissinger *et al.*, 2020). In this respect, the case of Airbnb stands out as a prime example. Marketed with the idea to enable individuals to offer spare space for rent, Airbnb is now a major player in the hospitality business around the world. The third and fourth chapters of this dissertation assess the positive and negative consequences that this extraordinary rise has for consumers, the hotel industry, and the rental market.

The last chapter of the dissertation analyzes the role of review and rating systems in peer-to-peer markets. Rating systems generate trust between users by allowing to learn about the trustworthiness of market participants. Thus far, one issue that has not received much attention is the potential consumer-harms related to online rating mechanisms (Bolton *et al.*, 2013; Luca and Zervas, 2016). The last chapter of this dissertation investigates whether price-setting can be used to strategically generate higher ratings and extract additional surplus.

The remainder of this introduction shortly summarizes each chapter. The goal is to describe the policy problems in more detail and to give a preliminary overview of the results.

## 1.2 Chapter Two: Learning from Data and Network Effects: The Example of Internet Search

Chapter two, which is joint work with Geza Sapi, analyzes the role of data for market power. Data are the central input for recommender systems, which are at the core of most business models monetizing on the accurate prediction of user preferences and choices. Recommender systems are used in various commercial applications: To suggest search results (Google Search, Yahoo!), products (Amazon, iOS App Store, Google Play), movies or songs (Netflix, Spotify), and the content of timelines in social media (Facebook, Instagram, TikTok, Twitter). The technology underlying recommender systems is often referred to as machine learning or artificial intelligence (Nilsson, 2014).

Machine learning algorithms, unlike classical algorithms, do not rely on a prespecified set of deterministic instructions. Instead, machine learning algorithms take a probabilistic approach to problem solving: Given some data and a problem, the algorithm learns which information in the data allows for solving the problem in the most reliable way. In some sense, the algorithm determines by itself the set of instructions to follow to best solve the problem. In fact, machine learning algorithms are often too complicated and too intricate for a human to understand their decision-making.

The notion that machine learning algorithms are essentially self-learning systems has raised concerns that the combination of massive data sets and machine learning leads to monopolized market structures (The Economist, 2017, 2018). Once a company achieves an edge in the amount of data it controls, a positive feedback loop sets in: Since more data imply higher quality through the self-learning technology, the company with more data will attract more customers, thereby increasing the amount of data it controls and so on... This narrative, also known as the *data feedback loop hypothesis*, is often put forward to explain the quasi-monopolistic market structures observed in many digital markets (Newman, 2014).

Among economists, the validity of the data-feedback loop hypothesis is heavily debated. One argument often put forward to argue against its validity is that learning from data displays diminishing returns to scale. This property of learning from data can be argued for theoretically: In statistics, estimators tend to display convergence properties, which imply large initial gains from data that fade off rapidly (Bajari *et al.*, 2019). While it is not clear whether this theoretical result from statistics carries over to machine learning, the diminishing returns to scale hypothesis appears to be validated by empirical results focusing on machine learning applications (He *et al.*, 2017; Yoganarasimhan, 2020; Bajari *et al.*, 2019; Claussen *et al.*, 2019). While not sufficient to rule out market power from data, the finding of non-increasing returns rules out one obvious mechanism rationalizing the data feedback loop hypothesis.

The second chapter of the dissertation investigates an alternative mechanism rationalizing the data feedback loop hypothesis. The analysis uses data from the Yahoo! search engine. We observe users entering search keywords and the associated search results. In particular, for each keyword, we observe how many users entered the keyword. More users lead to an improvement of search result quality because additional users imply more data collected on a keyword. If for the same number of users, one keyword improves faster in quality than another keyword, the search engine technology is more efficient for the faster-learning keyword. The analysis reveals that the speed of learning from additional users appears to be systematically related to the (average) amount of data the search engine collected on the users searching a keyword. This finding is consistent with data shifting the efficiency boundary of the technology outwards, which could be seen as a natural characteristic of a self-learning technology (Posner and Weyl, 2018).

The analysis underscores the importance of distinguishing between different sources of data generation. More users searching a keyword lead to data generated *across users*. The usage intensity of individual users instead leads to data generated on the level of single individuals: a user who visits the search engine frequently generates more individualized data. The results indicate that individualized user-data ameliorate the efficiency from learning across different users. In the machine learning literature, recommender systems relying on the combination of both data sources are known as collaborative filters (Adomavicius and Tuzhilin, 2005). From an economic perspective, one can view learning from data generated across users as a data-driven direct network effect: An individual user benefits from other users of the search engine through across user learning. Individual user data reinforce this direct network effect. This additional externality can be viewed as a *data-network effect*, since it captures network externalities stemming from user-data.

## 1.3 Chapter Three: Airbnb, Hotels, and Localized Competition

The third chapter of this dissertation, which is joint work with Kevin Tran, assesses the welfare effect from Airbnb's entry in the short-term accommodation market. Airbnb is one example of a peer-to-peer platform whose massive growth poses challenges for policy makers around the globe (Edelman and Geradin, 2015). In the context of the short-term accommodation market, Airbnb's presence is likely to affect consumers positively by offering additional accommodation choices. At the same time, it is also likely that Airbnb has a negative impact on the hotel industry because of the increased competitive pressure it exerts. To inform policy makers, it is important to understand whether the gains for consumers outweigh the losses for the hotel industry.

To quantify the welfare effects from Airbnb's entry in the accommodation market, we estimate a structural model of consumer demand and calibrate a model of hotel supply. Our analysis focuses on the accommodation industry in the city of Paris for the year 2017. We use three different datasets: For hotel demand, we use data provided by the French National Institute of Statistics and Economic Studies (INSEE). This data set, which is based on a monthly survey of French hotels, contains daily-level information on the number of rooms occupied for all hotels in the city of Paris. We complement the information on the occupancy of hotels with web-scraped hotel price data from the Booking.com price comparison platform. Daily-level information on the demand and the prices for individual Airbnb accommodations were provided by AirDNA, a company specialized in gathering data from short-term rental platforms.

From a methodological perspective, our main contribution is twofold: First, our model and the fine-grained nature of our data allow us to take into account the geographical dimension of competition within a city. Second, our supply-side model allows for capturing the pricing behavior of hotels in a very flexible and realistic manner. By taking into account the geographical dimension of competition, we can model the fact that close-by accommodations are closer substitutes than accommodations that are located further away from each other. This is a realistic model feature that is novel to a literature that mostly exploits variation across different cities (Zervas *et al.*, 2017; Farronato and Fradkin, 2018). On the supply-side, our model allows for capturing the steeply increasing pricing function that results from the capacity constraints that hotels face. This is important for capturing the price response of hotels to Airbnb's entry.

We perform a counterfactual analysis to simulate demand and the resulting price response of hotels if Airbnb were absent from the market. This allows for quantifying the welfare gains for consumers and the losses for hotels from Airbnb. We estimate that the average consumer surplus generated by Airbnb amounts to 4.2 million euros per night. This value takes into account the higher prices hotels would charge if Airbnb were not present in the market. The exact consumer surplus from Airbnb depends on the demand state: In periods of high demand, the consumer surplus can be as high as 8 million euro. Hotels incur an average revenue loss of 1.8 million per night. Thus, we find that Airbnb has a positive net welfare effect in the hospitality market.

## 1.4 Chapter Four: Airbnb and Rents: Evidence from Berlin

The fourth chapter, which is joint work with Tomaso Duso, Claus Michelsen and Kevin Tran, assesses the impact of Airbnb on prices for long-term rentals in the city of Berlin, Germany. The widespread use and adoption of Airbnb has raised concerns that long-term rents might increase as a consequence. Because Airbnb offers the possibility to generate revenue through short-term renting, this might reduce supply and, consequently, raise prices for long-term rentals. The potential adverse socio-economic consequences (Dudás *et al.*, 2017; Ferreri and Sanyal, 2018; Nieuwland and Van Melik, 2020) call for an exact quantification of the effect to inform the policy debate.

The analysis combines data on Airbnb supply and rental prices. For both Airbnb and rentals, we observe individual characteristics like the exact geographical location and the number of rooms. We augment these data with detailed information on neighborhood characteristics, such as socioeconomic information, pollution data, and noise data. In total, our dataset contains more than 170 variables.

Our strategy to identify the impact of Airbnb on long term rents exploits quasi-experimental variation in Airbnb supply. In 2016 and 2018 two laws took effect, each with the aim of curbing the supply of apartments on Airbnb. The first law, which took effect in May 2016, banned the misuse of apartment for commercial short-term renting. The second law, which took effect in August 2018, required hosts to display an official registration number on the Airbnb website.

In the analysis, we first document that both laws had a heterogeneous effect on Airbnb supply. While both laws led to a substantial and significant decrease in the number of full apartments listed on Airbnb, only the first law reduced the number of apartments rented out full-time. This indicates that the first law might have been more successful in increasing supply on the long term rental market: Landlords who rented out full-time on Airbnb likely had to substitute with long-term renting to achieve similar levels of occupancy.

For identification of the effect, we focus on three month time frames around both policy reforms. We employ a two-stage least square estimator with a post treatment dummy as instrument. To select among the high-dimensional set of covariates, we use the double-lasso estimator introduced by Belloni *et al.* (2012) and Chernozhukov *et al.* (2015). Our results indicate that both reforms reduce asked long-term rents shortly after their introduction. This is consistent with the hypothesis that Airbnb exerts upward pressure on the price of long-term rentals by reducing their supply.

For the first reform, we find stronger and more robust estimates, which is consistent with our hypothesis of the first reform being more successful in increasing supply of long-term rentals. We perform robustness checks focusing only on apartments rented out full-time and find stronger effects, which further corroborates the hypothesis. We also provide evidence that the reforms had a heterogeneous impact across the districts of Berlin.

Our results confirm previous findings in the literature documenting an increase in rents from Airbnb (Horn and Merante, 2017; Koster *et al.*, 2018; Garcia-López *et al.*, 2020; Barron *et al.*, 2020; Valentin, 2020). Compared to the existing literature, our granular data allows for assessing the impact of Airbnb on a more disaggregated geographical level and for directly controlling for rental specific characteristics. We find that one additional full apartment on Airbnb within a 250 meter radius increases asked rents by seven cents per square meter. A simple back of the envelope calculation indicates that Airbnb's presence leads to a 50 euro increase of the average asked monthly rent in Berlin.

## 1.5 Chapter Five: Strategic Pricing and Ratings

The last chapter of the dissertation analyzes the impact of prices on ratings on online platforms. Ratings are crucial to create trust between sellers and customers in the peer-to-peer economy. The economic literature devotes a lot of attention to the value of ratings for sellers. Several studies demonstrate the positive effect of ratings on sellers' revenues (Livingston, 2005; Luca, 2016; Resnick *et al.*, 2006). These results suggest that sellers have an incentive to influence ratings in order to boost revenue.

The chapter combines transaction and rating data from Airbnb to explore one channel through which sellers might influence ratings. The results show a robust negative correlation between the prices charged for an accommodation and the ratings it receives, which confirms the results of Luca and Reshef (2020). This finding indicates that travelers tend to factor prices into their overall assessment of a particular accommodation.

Based on this insight, I analyze if the data support the hypothesis that low entry prices confer a long run advantage. The analysis provides supporting evidence that hosts charging lower introductory prices benefit from higher initial ratings. This initial rating advantage seems to translate into higher revenues in the long run. The evidence suggests these higher revenues are achieved through gradual price increases. Interestingly, listings with low introductory prices maintain a rating advantage despite subsequent price increases. This indicates that high initial ratings might act as a reference point for subsequent ratings.

# Chapter 2

# Leaning from Data and Network Effects: The Example of Internet Search[1]

**Bibliographic information: Schaefer, M., Sapi G., 2020. Learning from Data and Network Effects: The Example of Internet Search, DIW Discussion Paper No. 1894.**

## 2.1 Introduction

The role of data in the success of firms is the subject of much debate. Data are often mentioned as a crucial input of production, so much so that academics and the press label data as "*the world's most valuable resource*" (The Economist, 2017). From a competition policy perspective the success of tech companies relying on data as one of their main inputs raises concerns that data might constitute a source of market power.[2] According to the "*data feedback loop*" hypothesis (Newman, 2014) there is a self-reinforcing cycle between the success of firms and the amount of data they control. This hypothesis is often put forward to explain the rise of dominant players in data driven industries.[3]

While the rise of data-driven superstar firms in digital industries such as Amazon, Google, Facebook, Netflix and Spotify lends some credit to the hypothesis that data might constitute a source of market power, independent empirical evidence systematically examining the impact of

[1]This chapter is joint work with Geza Sapi. An earlier version of this paper has been published under the title "The Effect of Big Data on Recommendation Quality: The Example of Internet Search" (Schaefer *et al.*, 2018), which was additionally coauthored by Szabolcs Lorincz. We are extremely grateful to the many comments and suggestions provided by Tomaso Duso and Hannes Ullrich.

[2]Depriving rivals of data was an important pillar of the European antitrust proceedings against Google, resulting in a record fine of 4.3 billion euro (European Commission, 2018, recitals 111, 114, 458, 514, 739, 860(3), 1318, 1348)

[3] The European Commission announced its plan to enforce opening up the data of dominant tech companies to rivals. It argued that "*the high degree of market power resulting from the 'data advantage' can enable large players to set the rules on the platform and unilaterally impose conditions for the access of data.*" (European Commission, 2020)

data on the performance of firms remains scarce. The handful of existing studies on the topic suggest that data are essentially like any other ordinary input: All of these studies report evidence for diminishing returns to scale from additional data (He *et al.*, 2017; Yoganarasimhan, 2020; Claussen *et al.*, 2019; Bajari *et al.*, 2019). If there were diminishing returns to scale in data, it may be relatively easy for potential entrants to obtain sufficient data to be viable. The finding of these studies therefore cast some doubt on the popular narrative of data being a *special input* and the root cause of competitive advantage.

In this paper we provide empirical evidence that additional data matter and document how data improves the quality of internet search results. Our findings indicate that data do not simply constitute an input into a static production technology: instead, they simultaneously shift the efficiency boundary of the search result matching technology outwards. While our empirical evidence is consistent with a technology that displays diminishing returns to scale, the improvement of efficiency with larger amount of data provides a novel rationale for an inherent competitive advantage from "*big data*".

We analyze search traffic data from Yahoo!. We observe users entering keywords in the search bar of the search engine and follow their subsequent interaction with the search results. The search engine collects the logs of users' clicks on the search result page shown in response to these queries. The data obtained are valuable because they allow the search engine to learn from the observed click behaviour and improve the quality of search results.

Over time, data on the users' action accumulate along two different dimensions. First, as more users enter a specific keyword, the amount of data collected on the feedback for that keyword increases. We call this the *keyword dimension* of data accumulation. Generally, data on a specific search keyword are collected across different users. The amount of data collected in the keyword dimension is therefore directly related to the number of individuals searching that keyword. Second, the more often the search engine observes the clicks of individual users, the more information it can collect on the individual preferences and interests of users. We refer to this as the *user dimension* of data accumulation.

Both dimensions are relevant for the predictive performance of the search engine. More data in the keyword dimension allow the search engine to determine common preferences across users. A user being the first to enter a keyword will generally be confronted with results of a lower quality than a user entering the same keyword at a later point in time. This is because, over time, the feedback provided by users allows for determining a general quality ranking of search results. Learning in the keyword dimension can be viewed as a data-enabled direct network effect: Additional users will generally increase the quality experienced by a specific user through learning from data in the keyword dimension.

The data collected on the user dimension allow the search engine to derive preferences specific to the user. In conjunction with the information that the search engine obtains on other users, data on the user dimension allow the search engine to determine user profiles. Users with similar profiles, revealed through overlapping preferences in the past, will also be more likely to have overlapping preferences in the future. When confronted with the task of finding relevant search results for a specific user searching a keyword, the search engine builds its prediction based on the feedback obtained from similar users who previously searched the same keyword.

In this paper, we highlight the importance of the interaction between the user and keyword dimensions of data for the ability of search engines to learn. Our main result is that the improvement of search result quality through additional data in the keyword dimension is positively affected by the amount of data collected in the user dimension. Keywords that are repeatedly searched by different users improve faster in quality when the search engine has more information, i.e. data, on the users searching that keyword. Learning from additional data in the keyword dimension becomes more efficient with additional data in the user dimension. Hence, search engines with longer search histories on their users learn faster as their user base grows.

Intuitively, the result can be understood in the following way. The value of the feedback that a user provides is determined by the amount of information the search engine has on the user. More information on the user makes the feedback she provides more valuable because the search engine can relate the feedback to a more detailed user profile. More information on users searching a specific keyword allows the search engine to tailor search results to specific user profiles more efficiently. If the search engine has, on average, more data on the searchers that provide feedback on a specific keyword, this will lead to a faster quality increase as a function of the data collected across different users searching that keyword. More data in the user dimension allow the search engine to determine more rapidly what a specific user wishes to see when searching a specific keyword. Without personal information, the search engine cannot learn which type of user prefers which type of search results for a specific keyword. It is only by combining individual data across different users that the search engine can learn which information in the user profiles is relevant to individualize search results for a keyword. The more information the search engine has on the user profiles, the more likely it is that it will pick up the relevant information necessary to improve the quality of search results for specific user types.

Our results provide evidence that data-driven services benefit from an externality that goes beyond usual direct network effects. Learning in the keyword dimension is comparable to direct network effects: More users generate more data on keywords which benefits other users of the search engine. Additionally, our analysis suggests that these direct network effects are reinforced by data collected in the user-dimension. Augmented by user-specific data, learning in the keyword dimension occurs faster and, thus, more efficiently.

This additional externality is unique to data-driven industries and confer an additional competitive advantage to the incumbent firm. To understand the implications of our results consider two firms A and B with the same number of users: Under conventional direct network effects, both firms should be equally competitive because they benefit from the same size in user base. However, our results suggests that if one firms has more information on its users, it will be more competitive by virtue of the additional externality that we reveal, which can be viewed as an additional *data network effect*.

From a competition policy perspective, our results are consistent with the hypothesis that a firm with a larger database has a competitive advantage and that lack of data constitutes a barrier to entry. Our results therefore call for a consideration of database sizes in merger decisions between firms that rely heavily on data for their business models. Our findings suggest that the ability to combine data on users across different services might be extremely valuable.

Our paper is related to a nascent strand of literature dealing with the impact of *big data* on firm performance. Several contributions approach the topic of scale economies in data from a policy perspective: Lambrecht and Tucker (2015), Sokol and Comerford (2015), and Tucker (2019) argue that the era of digitization poses no special challenge for antitrust authorities and that network effects from data accumulation should be expected to be weak. Newman (2014) and Grunes and Stucke (2015), on the other hand, argue that data can play an important role for firms in securing competitive advantages over rivals and call for a reorientation of antitrust policy to better account for the role of data as a barrier to entry. Schepp and Wambach (2015) submit that current competition law should be flexible enough to address the new challenges posed by digitization but emphasize the role of data in understanding dynamics in digital marketplaces. Argenton and Prüfer (2012), Prüfer and Schottmüller (2017) and Hagiu and Wright (2020) model competition in data-driven markets. Economies of scale due to nonrivalry of data and associated welfare effects have been analyzed in Jones and Tonetti (2020). Casadesus-Masanell and Hervas-Drane (2015) analyze the implication of consumer privacy for competition.

Bajari *et al.* (2019) provide theoretical support for diminishing returns from data and empirically analyze the impact of data on the predictive performance of Amazon's retail forecast system. Their findings are consistent with diminishing returns from repeatedly observing a product in the forecast system. Additionally, Bajari *et al.* (2019) investigate to what extent observing additional products in the same product category improves performance of the algorithm and find no noticeable effect. As a result, they conclude that economies of scope from data are weak. In contrast, He *et al.* (2017) find indication for economies of scope in the context of search engine data.

Our findings are related to the notion of economies of scope according to which algorithms benefit from increased data-variety in addition to sheer data-quantity. Our results reveal that combining different sources of data (user-specific data with data gleaned across users) benefits the firm by further increasing the efficiency of the matching technology. However, the mechanism of action we propose is also distinct from economies of scope in that it builds on synergies between user-specific data and the size of the user base. Our insights speak to the notion that artificial intelligence continuously improves with data. This notion is also put forward in Posner and Weyl (2018), who argue that returns from data in the machine learning context follow a different paradigm than in classical statistics. Our findings provide empirical support for their view, according to which the true value of data in the machine learning context can only be assessed by considering the "*overall learning of the system*" (Posner and Weyl, 2018, p. 227).

Our research is further related to Chiou and Tucker (2017) who use an exogenous policy change in the data retention policy as an identification strategy to analyze returns from data. Chiou and Tucker (2017) find no indication that reducing the retention time of user specific information affects search result quality. Claussen *et al.* (2019) and Yoganarasimhan (2020) document the important role of personalized data for the predictive performance of algorithms. Both studies find evidence for diminishing returns to additional data collected on users. We add to this literature by exploring the simultaneous interaction between the user and keyword dimensions of data and by proposing a new mechanism whereby data on individual users reinforce the learning from data on keywords. Our paper is also broadly related to Decarolis and Rovigatti (2019) and Decarolis *et al.* (2020) who study competition in the online advertising market for search engines.

The paper proceeds as follows: In Section 2.2, we present the data and define the quality measure we use for analysis. Section 2.3 outlines our empirical strategy to assess network effects in data as the search engine learns from more data. Section 2.4 presents the results of our empirical analysis, which consists of two parts. The first part presents evidence for the search engine learning from data by taking a long run perspective. Using a proxy variable technique, we show that keywords with more data on the user dimension prior to the period our sample was constructed achieve a higher observed quality level in our sample. The second part focuses on the observed quality evolution of search results returned to keywords during our sample period. We show that the quality of search results improves with more data accumulating on the keyword. The quality improves faster when the searches are entered by users on whom the search engine already gathered more data. Section 2.5 discusses identification using a stylized model of statistical learning. Section 2.6 concludes.

## 2.2 The Data

The data we use stem from Yahoo! and contain fully anonymized search logs spanning a period of 32 days from July 1, 2010 till August 1, 2010, inclusive. An observation in our database contains a keyword identifier, a cookie identifier essentially corresponding to the computer on which the search was conducted, the precise time the keyword was entered in the search bar, the ordered list of the top ten organic result Uniform Resource Locators (URLs) and the sequence of clicks performed by the user. In total, we have approximately 80 million observations from 29 million users (identified by the cookie) searching for 67 thousand different keywords. As the search keywords are anonymized, we do not observe the number of words in the query. We however observe when the same combination of keywords was entered as these search terms have the same identifier.

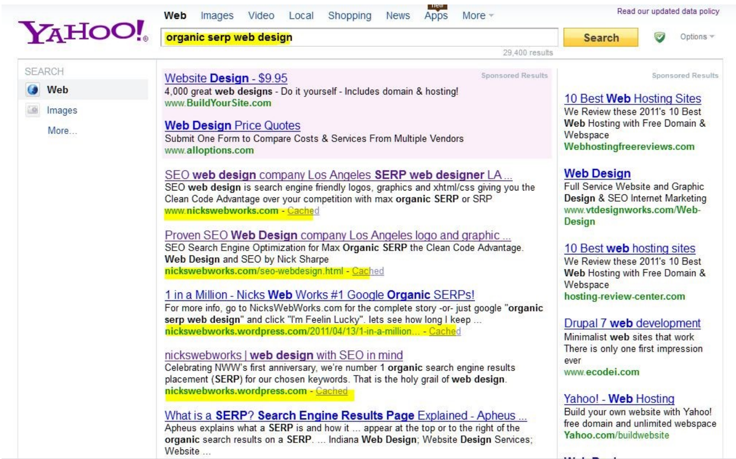



**Figure 2.1:** Search result layout at Yahoo!, 2011

Figure 2.1 illustrates the structure of the typical search result page at Yahoo! at the time the data were collected. The search keyword, highlighted in yellow next to the Yahoo! logo, is the sequence of characters the user enters in the search bar in her quest for information. Our analysis focuses on the quality of the organic search results URLs, which are highlighted in yellow in the search result list. Paid advertisements are displayed on the north and east edges of the result list. This general layout of the search result page remained largely intact up to today, with search engines typically devoting the top and east edges of the result page to ads, and present up to 10 organic search result URLs in the middle area.

### 2.2.1 Measuring Search Result Quality

For each search session we observe a log that records the actions of the user on the result page. This usually consists of a series of clicks on the retrieved search result links, possibly returning from those URLs to the search result page, until the user clicks on a search result link from where she does not browse back to the result page to continue her search. If the user does not return to the result page to continue her search, the log ends. To measure the quality of search results to a specific keyword, we create a binary variable of search result quality based on the users' actions taking the values *good* (encoded as 1) or *bad* (encoded as 0).

We deem the search result quality as *good* if the last recorded click in a log occurs on the top displayed organic URL. This measure takes into account the fact that users have a natural tendency to perform their first click on the top displayed URL. If a user performs her first click on the top displayed URL and subsequently returns to the result page to choose a URL further down the result list, the search result quality is encoded as *bad*.

Based on this criterion, we calculate the click through rate as the fraction of searches ending on the first URL. The click through rate on the first URL for keyword $i$ is defined as:

$$ctr_i^{\{1\}} = \frac{\sum\limits_{s_i \in S_i} \mathbb{1}\{lcp_{is} = 1\}}{\sum\limits_{s_i \in S_i} \mathbb{1}\{lcp_{is} \neq 0\}} \tag{2.1}$$

Where $\mathbb{1}$ denotes the indicator function and $lcp_{is}$ the last click position recorded for search $s_i$. $S_i$ is the set of searches considered in the computation of the click through rate for keyword $i$. For example, if the click through rate is computed for a particular day, then $S_i$ is the set of all the searches on that particular day. Finally, $lcp_{is} = 0$ denotes a final click on an advertisement URL, which we ignore in the computation of the click through rate. Advertisement URLs are paid content that might be displayed even if the content is not relevant to the user. Because we want to measure the performance of the search engine in its ability to find relevant content for searches where the user was interested in organic content, we decided to ignore clicks on advertisement URLs in the main quality measure. Our results remain qualitatively unaffected if we consider alternative click based quality measures or if we include ads in the analysis (see Appendix 2.7.1.3 and Appendix 2.7.1.4).

Click through rates on the top URLs are an intuitive measure for search result quality, whose variations are widely used in the information retrieval literature (Joachims, 2002). Users expect the most relevant search result be displayed at the most prominent position on the search result page. A user clicking on top-displayed URLs and not returning to the search result page means she was satisfied with the information found under that link.

In the Appendix we assess the robustness of our results when using an editorial quality measure. In particular, our data set also contains 659,000 query-URL pairs with editorial relevance judgments Yahoo! collected from human experts. The correlation coefficients between the $ctr^1$ measure and two editorial quality measures that we consider are 0.45 and 0.55, and hence clearly show positive correlation of moderate magnitude. The editorial quality measure is explained and discussed in depth in Appendix 2.7.1.1. We present the results of our main analysis when using the editorial quality measure in Appendix 2.7.1.2.

## 2.3 Empirical Strategy

### 2.3.1 Measuring Learning from Data

In this subsection we discuss the variables used to study the search engine's ability to learn from data and improve the quality of search results. We focus on two different dimensions of data: First, the *keyword dimension*, which captures the amount of data that accumulates through searches performed on a particular keyword. The number of users determines how much training data the search engine has for a given keyword. In the *keyword dimension*, learning from data is akin to direct network effects: with more users searching on the same keyword, the search engine has more training data available to solve the task to predict the most relevant search results to display to that particular keyword.

The second dimension captures the amount of data that the search engine collected on the users searching a specific keyword. The value of user-specific data is widely recognized by companies relying heavily on predictive technologies, such as Netflix, Amazon and Google.[4] For internet search, it is common wisdom that data on user search histories are useful because they allow fine-tuning search results for the individual preference of each searcher. In this article we emphasize that the importance of data on the *user dimension* goes beyond the fact that they allow fine tuning search results for a particular user: We show that more data on the *user dimension* - longer observed search histories for users - allow the search engine to learn faster from more data on various keywords. Data in the *user dimension* therefore acts as a catalyst for learning from data in the *keyword dimension.* They can amplify network effects stemming from more users and are a key yet so far underemphasized source of competitive advantage and market power.

Our results are consistent with the following learning mechanism that drive network effects based on data: When a user searches for a particular keyword, the search engine engages in user profiling to establish similarities between the current user and past users who entered the same keyword. Search results that proved relevant to past users with a similar profile are also more likely to be relevant for the current user. Establishing similarities between the current user and past users is facilitated by the amount of data the search engine collects about users. Intuitively, the more data the search engine collects on users, the more likely it is to find overlap in their browsing behaviour

With more data collected in the *keyword dimension*, the search engine learns which search results are relevant for similar user profiles by deducing which information in the overlapping browsing behaviour of user is informative about their preferences for the keyword. Because more data on the users makes it easier to elicit the relevant characteristics to determine relevant search results for specific user profiles, the search engine is able to retrieve more relevant search results when more data on the users are available.

In the machine learning literature, the idea of training algorithms for prediction tasks based on the overlap of past user preferences is known as collaborative filtering.[5] It appears intuitive that this method requires both training data in the *keyword dimension* and data in the *user dimension.* A user being the first to search a keyword may likely be confronted with relatively poor search results because the search engine has no previous training data on the keyword. Similarly, a lack of data on the users does not allow the search engine to tailor results to user profiles. We argue here that these two dimensions of data are not independent from each other: the amount of data in the *user dimension* increases the efficiency of learning in the *keyword dimension.* This in turn has serious implications for economies of scale, and consequently for competition between search engines. To the best of our knowledge, we are the first to systematically study the interaction between the *user* and *keyword dimensions* of data. We now introduce the variables capturing both data dimensions.

---

[4]Netflix explains on its website the use of person-related information to suggest movies: "*Personalization is one of the pillars of Netflix because it allows each member to have a different view of our content that adapts to their interests and can help expand their interests over time. It enables us to not have just one Netflix product but hundreds of millions of products: one for each member profile*" (Netflix, 2020).

[5]See Adomavicius and Tuzhilin (2005) for an overview of the different architecture types of recommender systems.

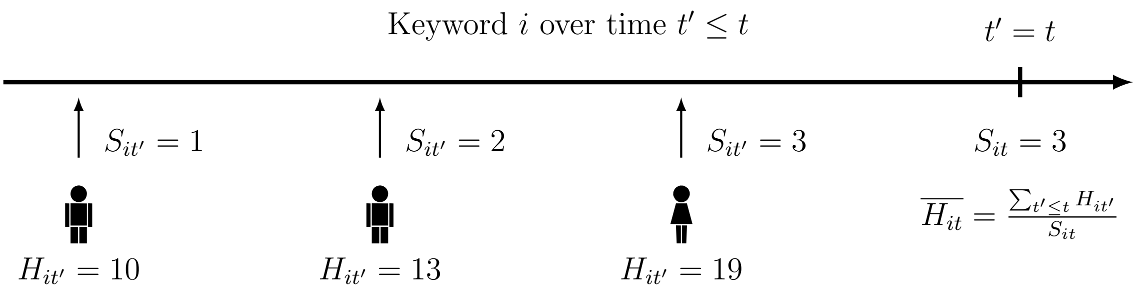



**Figure 2.2:** Variable description

- $S_{it}$: **The cumulative number of searches** for keyword $i$ at time $t$. It describes the size of training sample for keyword $i$ at time $t$.
- $\overline{H_{it}}$: **The average user history** for keyword $i$. It captures the average amount of data the search engine had on the users that entered the keyword in the search bar of the search engine until time $t$. Denote by $H_{it'}$ the length of the search history of the user querying keyword $i$ at time $t' \leq t$, then $\overline{H_{it}} = \frac{\sum_{t' \leq t} H_{it'}}{S_{it}}$.

Figure 2.2 illustrates how both variables are computed at a given point in time $t$. The average user history is a natural measure for the amount of data the search engine collected on the users who searched for a specific keyword. Throughout the analysis, we track search result quality as a function of $S_{it}$, a measure for the amount of training data available for keyword $i$. We explore a novel externality in learning from data that provides a new perspective on the value of data. Data on individual users are particularly valuable, because learning from data along the *keyword dimension* becomes faster when the average amount of data collected in the *user dimension*, captured by $\overline{H_{it}}$, increases.

### 2.3.2 Data Generating Process

Before turning to the analysis of search result quality and data volume we pause to discuss the nature of our sample in more detail.

Our dataset is a snapshot of searches observed through a one month period. We do not know how often the search engine had observed the sampled keywords and users prior to the sample period. For example, in theory it is possible that the keywords we observe often are only popular during the month of our sample, but had never been searched upon before. This may be the case for example for news related searches. We do not directly observe the overall amount of data the search engine had collected on keywords and users before our sample started, because we only observe a snapshot of the overall search traffic.

With our data spanning over a one-month period, the variables that we observe constitute the monthly counterpart of the true variables we are interested in, namely $S_{it}$ and $\overline{H_{it}}$. We denote the observed counterparts $s_i$ and $\overline{h_i}$ for which we drop the index $t$ since we focus on a fixed time span of one month. Intuitively, keywords with more searches and longer user histories during the month of our data are likely to also have experienced more searches and longer user histories previously. In the further analysis we will restrict attention to keywords for which we observe approximately the same number of searches every day. The vast majority of the keywords in the sample fall into this category, as relatively few keywords experienced rapid jumps or drops in the number of searches over time in our sample.

Focusing on keywords with a relatively stable search activity over time is useful: since these keywords see a fairly constant number of searches every day during the period we observe them, it appears reasonable to assume that these keywords had seen a similar search pattern also prior to our sample period. If that is the case, we can compare keywords that received more and less searches in our sample and assume that this also held in the period before the sample started. In other words, the monthly counterparts $s_i$ and $\overline{h_i}$ constitute good proxy measures for $S_{it}$ and $\overline{H_{it}}$.

This would be much less likely to hold for queries that experience a rapid boost in searches in the sample: such queries may for example relate to news, where we would not know how often the keyword had been entered prior to our sample period. For example, consider a rapidly booming query observed 50 times, all on a single day and never otherwise in the sample, and another query observed once on each of the 32 days of our sample. While in the sample it appears that the booming query was observed more often (50 times), it is possible and even likely that the search engine has more observations on the query that occurs less often (32 times) in the sample: that query was most likely entered regularly also prior to our sample.[6]

To select the keywords with relatively stable search patterns over the days of our sample, we calculate for each observed keyword the number of searches per day. Take first the limiting case of a query that does not experience popularity changes during the sample period and receives the same number of searches each day. Then this query accumulate each day 100/32 percent of its total observed monthly searches. We will refer to the special case with the number of daily searches uniformly distributed as *even accumulation criterion.* We define tolerance levels that determine the maximum percentage point deviation a keyword is allowed to have on one particular day. For instance, a tolerance level of ten percentage points indicates that on each day of the sample, a keyword is not allowed to accumulate more than $(100/32+10)$ percent of its total searches in order to be retained in the sample.

Intuitively, the narrower the tolerance interval, the better the quality of $s_{it}$ and $\overline{h_{it}}$ as proxy variables. The largest tolerance level we consider is 50 percentage points, such that a keyword is allowed accumulate up to $100/32 + 50 \approx 53$ percent of its total searches in one day without being dropped. The narrowest tolerance level is ten percentage points, meaning that a keyword is allowed to accumulate up to 13 percent of its total searches in one day without being dropped.[7]

Importantly, while focusing attention to keywords that see relatively similar number of searches every day allows us to reduce the noise in our data, this is not crucial for our analysis. We obtain the same qualitative effects when using the full sample as well.

In subsection 2.4.1 we will show that keywords with longer user histories and more searches previous to our sample have higher quality search results.

## 2.4 Results

Our empirical analysis consists of two parts. In the first part, we take a long run perspective on the data. We take into account that the search engine may have gathered data on the keywords we

[6]In Appendix 2.7.2, we discuss in more depth the detailed assumptions under which it is reasonable to focus on constantly searched queries.

[7]Nearly 90 percent of the keywords display a search pattern over time that falls within the narrowest tolerance level we consider.

observe already before the sample period. Depending on the number of searches and the average user history before our sample period, keywords should differ in the quality level we observe in our sample. If the search engine learns faster with additional data in the *user dimension*, keywords with a longer average user history in the past should, *ceteris paribus*, experience a higher quality level.

We first discuss long term trends, looking at the search result level of queries in our sample as a function of the proxied number of searches and average user history prior to the sample period. Focusing on keywords with a relatively stable number of searches over time helps better isolate how additional data drives learning.

In the second part, we analyze the quality evolution of keywords over the sample period. Building on the insights from the first part, we take into account the unobserved number of searches and average user history prior to the sample period. If the search engine learns faster with additional data in the *user dimension*, keywords with a longer average user history should, *ceteris paribus*, experience a larger increase in quality. We confirm that this is indeed the case. The learning observed within the sample period is consistent with the long run dynamics documented in the first part of the analysis.

### 2.4.1 Long Run Analysis

In this subsection we study the quality level of queries in our sample as a function of the number of searches and the average user history previous to the sample period. If search result quality increases faster with more data in the *user dimension*, we would expect that search results for keywords with a longer average user history prior to the sample period lead to higher quality search results when we observe them. If we could access the entire search history for each keyword and each user we observe in the sample, we would directly estimate the following equation:

$$ctr_i^1 = f(S_{i\underline{t}}, \overline{H_{i\underline{t}}}) + \epsilon_i \tag{2.2}$$

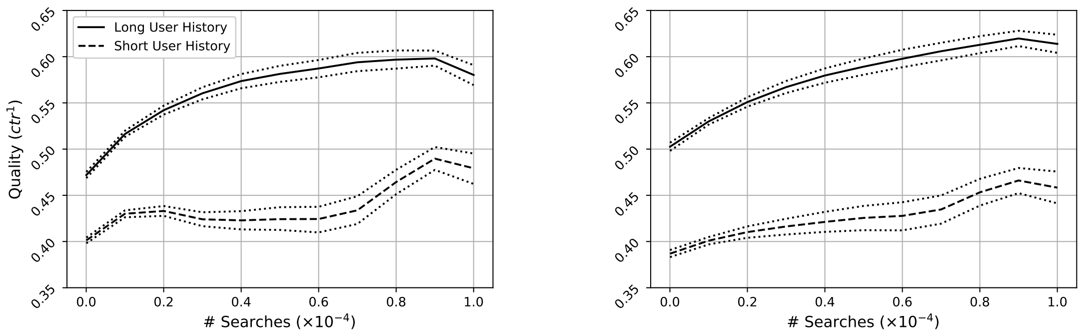
Where $ctr_i^1$ is the click through rate on the first URL for keyword $i$ based on all the searches we observe. $S_{i\underline{t}}$ the number of searches previous to the sample period and $\overline{H_{i\underline{t}}}$ the average user history previous to the sample. $\epsilon_i$ is the error term. Since we do not observe $S_{i\underline{t}}$ and $\overline{H_{i\underline{t}}}$, we estimate:

$$ctr_i^1 = f(s_i, \overline{h_i}) + \epsilon_i \tag{2.3}$$

Where $s_i$ and $\overline{h_i}$ denote the number of searches and the average user history we observe for keyword $i$ during our sample period.

Throughout this analysis, we assume that larger values of $s_i$ and $\overline{h_i}$ induce larger average values of $S_{i\underline{t}}$ and $\overline{H_{i\underline{t}}}$. Thus, we assume that (i) keywords for which we observe more searches during the sample were on average searched more often before the sample period, and, (ii) keywords with a longer average user history during the sample have, on average, a longer user history previous to the sample period.

Estimation is performed by local linear regression on a grid $s \times h$, where $s = \{0, 0.1, 0.2, .., 1\}$ and $h = \{Q_1, Q_3\}$. The values of $s$ denote number of searches in tens of thousand. $Q_1$ and $Q_3$




**(a)** Entire sample of keywords **(b)** Keywords accumulating evenly

**Figure 2.3:** Average search result quality of keywords as a function of the total number of searches and the average user history, with 95 % confidence intervals.

denote the lower and upper quartiles of the distribution of the average user histories.[8][9]

Figure 2.3 shows the results of our analysis. The solid line in Figures 2.3a and 2.3b correspond to keywords with longer user histories, taken at the upper quartile ($h = Q_3$) of the user history distribution. The dashed line depicts keywords with short user histories, which we set at the lower quartile ($h = Q_1$) of the distribution of user histories. In Figure 2.3a, estimation was performed using all the keywords in the sample. In Figure 2.3b keywords that deviate no more than ten percentage points from the even accumulation criterion (100/32 percent of searches per day) were considered. According to the discussion in subsection 2.3.2, the results in Figure 2.3b are more informative about the true relationship between the quality of keywords and $S_{i\underline{t}}$ and $\overline{H_{i\underline{t}}}$.[10]

The solid lines in Figure 2.3 map out the quality level as a function of the number of times the keyword was searched upon, $s$, for keywords with a long average user history. The dashed lines map out the quality level for keywords with a short average user history.[11] Note that when we focus on keywords that accumulate searches more evenly over time in Figure 2.3b, the observed pattern for keywords with a short average user history normalizes.

Figure 2.4 shows the difference in the search result quality level between keywords with a long and short average user history for a given number of searches, $s$. The black line in Figure 2.4 shows the difference for the results in Figure 2.3a. The line with the lightest grayscale shows the difference for the results in Figure 2.3b, when only those keywords were considered that accumulate searches over time most evenly. The other lines show the differences for intermediate tolerance levels with a lighter grayscale indicating a stricter tolerance level. Figure 2.4 reveals that the measured impact of the average user history increases as we gradually drop keywords that accumulate searches less evenly over time: Gradually reducing the measurement error in the proxy variables also increases the measured impact of the average user history. This is in line with what we expect to see when

[8] The number of searches we observe is truncated at $10,000$. The lower quartile is equal to 2.79, the upper quartile is equal to 4.78. These numbers might appear low but can be explained by the fact that we observe a random sample and that users are observed only three times on average in our sample.

[9] In Appendix 2.7.3, we present the details on the estimation method.

[10] The reader is referred to Appendix 2.7.2 for an in depth discussion of the relationship between $s_i$ and $\overline{h_i}$ and the unobserved $S_{i\underline{t}}$ and $\overline{H_{i\underline{t}}}$.

[11] The quartiles of the average user history distributions are determined based on all the keywords in the sample, i.e. they do not vary across the different populations of keywords considered.

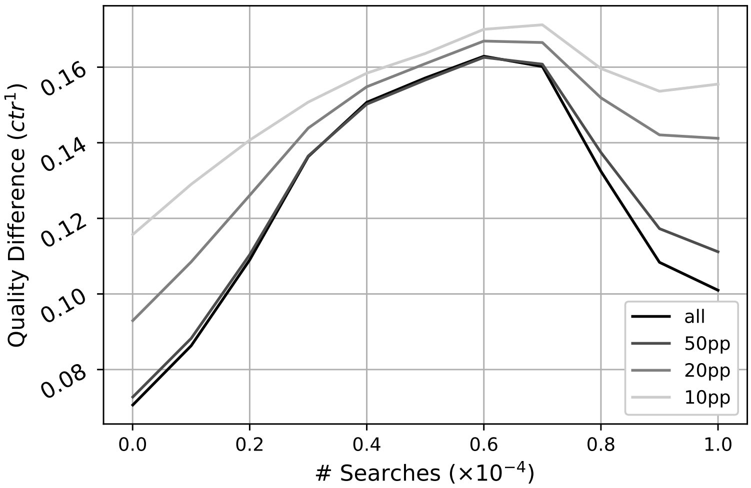



**Figure 2.4:** Impact of a longer average user history on the observed quality level. Different tolerance levels for deviation from even accumulation criterion are considered.

the relationship between the average user history and the quality of search results was causal and we increase the quality of the proxy variables.

From Figure 2.4, we also see that the difference in quality between keywords with a long and short average user history tends to increase with additional searches on the keywords. This divergence in quality levels is in line with the hypothesis of more efficient learning from additional searches through a longer average user history. The measured divergence will be at the core of Section 2.5, where we will analyze whether potential sources of confoundedness might rationalize the documented divergence when a longer average user history does not lead to faster learning.

The findings of this subsection can be summarized as follows: First, our data are consistent with diminishing returns to scale in the total number of searches, $S$. Second, we find that more data on the individual users amplify network effects: learning from the *keyword dimension* of data ($S$) becomes faster as more data on the *user dimension* ($H$) accumulates. Third, the relationship between search result quality and data in the *user dimension* becomes stronger when we focus on keywords that accumulate searches more evenly over time. Fourth, we observe a divergence in quality between keywords with a long and short average user history as more data in the *keyword dimension* ($S$) becomes available.

These results are consistent with the claim that data may constitute a source of market power. A search engine with more data on the individual search histories of its users will always outlearn the rival on new keywords, and any keyword that the two search engine had observed equal times previously. The following analysis of the learning in the sample confirms this finding.

### 2.4.2 Quality Evolution within the Sample Period

We now turn to the analysis of how search result quality changed during the 32 days of our sample period. We again focus on changes in search result quality as a result of more data accumulating in the *keyword* and *user dimensions*, respectively $s_i$ and $\overline{h_i}$. Consistent findings with the results from subsection 2.4.1 would imply a positive concave relationship between the quality evolution of keywords and $s_i$. Furthermore, for the same value of $s_i$, we would expect the quality increase to be more pronounced for keywords with a larger average user history, $\overline{h_i}$.

Subsection 2.4.1 reveals a strong relationship between the unobserved $S_{i\underline{t}}$ and $\overline{H_{i\underline{t}}}$ and the quality level of a keyword during the sample period. It is also to be expected that $S_{i\underline{t}}$ and $\overline{H_{i\underline{t}}}$ will strongly impact the quality evolution we observe for keywords during the sample period. For instance, in the presence of diminishing returns to scale, larger $S_{i\underline{t}}$ should, *ceteris paribus*, reduce the measured quality evolution of keywords.

The results from subsection 2.4.1 suggest that controlling for the quality level that keywords reached should help to control for $S_{i\underline{t}}$ and $\overline{H_{i\underline{t}}}$. We therefore estimate the following equation:

$$\Delta ctr_i^1 = f(s_i, \overline{h_i}, ictr_i^1) + \epsilon_i \tag{2.4}$$

$\Delta ctr_i^1$ denotes the difference in the click through rate on the first URL between the first and last 100 searches that we observe for a keyword. $ictr_i^1$ denotes the click through rate on the first URL for the first 100 searches that we observe for a keyword.[12] $ictr^1$ is an intuitive measure for the quality a keyword reached at the beginning of our sample period, which helps to account for differences in the unobserved $S_{i\underline{t}}$ and $\overline{H_{i\underline{t}}}$.

Estimation is performed by local linear regression on the grid $ictr \times s \times h$.[13] The values of $ictr$ are given by $\{0.25, 0.5, 0.75\}$ and the values of $s$ by $\{0.02, 0.1, 0.2, ..1\}$. The values of $h$ are identical to the ones chosen is subsection 2.4.1. Figure 2.5 shows the result of estimating Equation 2.4 on the grid $ictr \times s \times h$. Each panel stands for a different initial quality level. The solid lines within each panel map $\hat{f}$ as a function of $s$ for keywords with a long average user history, the dashed lines for keywords with a short average user history.

From the left panel of Figure 2.5, we can read that keywords with $ictr^1 = 25\%$, $h = Q_3$ and $s = 10,000$ experience an average increase of eight percentage points in the click through rate on the first URL. Keywords with $h = Q_1$ and otherwise identical parameters experience a quality increase of only 4.8 percentage points.

Data on the *user dimension* amplifies learning from data on the *keyword dimension*: Keywords with longer average user history, *ceteris paribus*, display a larger quality increase than keywords with a shorter average user history.[14] Consistent with diminishing returns to scale from additional searches on a keyword, we observe a concave pattern between the measured quality evolution and the number of searches, $s$. Furthermore, the average quality increase diminishes *ceteris paribus* with a larger initial quality level, which is also in line with diminishing returns in $S$, the number of searches on a keyword.[15]

[12] Note that by construction estimating Equation 2.4 requires us to focus on keywords with at least 200 searches during the period of our sample. Estimating Equation 2.4 for all the keywords in our sample is frustrated by regression to the mean. Regression to the mean arises in any kind of analysis where observations are classified based on a noisy measure of the initial outcome. The regression to the mean phenomenon and its impact on our analysis is discussed in more detail in Appendix 2.7.4.

[13] Details on the estimation method and bandwidth selection are given in Appendix 2.7.3.

[14] For estimation, we removed all keywords that accumulate more than 50% of the total searches during the sample period within a single day. Further narrowing the tolerance level of the even accumulation criterion does not significantly impact the results.

[15] The negative quality evolution measured for keywords with $ictr^1 = 0.75$ and $h = Q_1$ (dashed line in lower right panel) might be due to the fact that those keywords experience an increase in the click through rate on URLs positioned below the first URL. Thus, clicks on the URLs below the first one might cannibalize clicks on the first URL. In Appendix 2.7.1.3, we analyze the quality

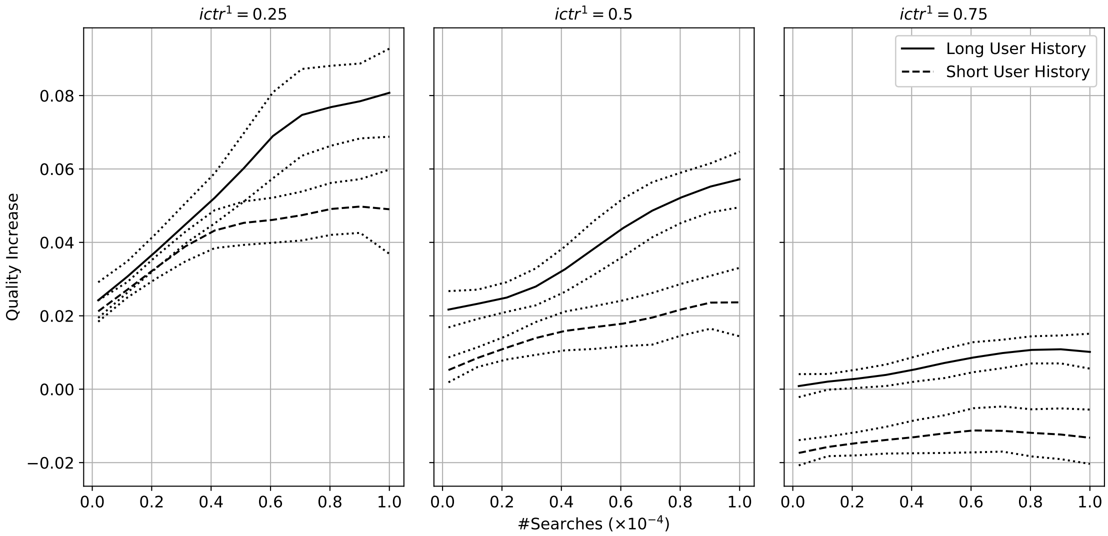



**Figure 2.5:** Average quality increase of keywords with 95% confidence intervals. Each panel stands for a different initial quality level.

In summary, four key results emerge from this analysis on how data matters for search result quality. First, search engines learn from data on user response to search results shown in response to the same keyword. Search result quality improves for queries that are searched upon more. Second, the personal dimension of data matters for learning. Data in the *user dimension* enable faster learning from data in the *keyword dimension*. Third, the learning does not fade off fast. We observe quality improvements even at the most searched upon keywords. Fourth, personal data matters even for new queries. This last point is particularly important for the question whether data may constitute an entry barrier. It implies that between two hypothetical search engines, even if they had an equal number of users, the new entrant will be at a disadvantage and increasingly so: the established firm will have observed longer search histories of its users and will be able to provide higher quality search results even to queries that are new to both search engines.

The findings from this subsection match our expectations from the analysis of the long run dynamics. The fact that keywords with a higher quality level improve less in quality is in line with the pattern of diminishing returns to scale from $S_t$, which we found in subsection 2.4.1. Additionally, we find larger quality increase for keywords with a longer average user history. This indicates that data on the *user dimension* allows the search engine to learn faster from data in the *keyword dimension*. This effect explains the marked quality differences between keywords with a long and short average user history in subsection 2.4.1.

## 2.5 Identification

Sections 2.4.1 and 2.4.2 showed that data in the *user dimension* enhance learning from data in the *keyword dimension*. We argued that more data may constitute a source of market power, as more data in the *user dimension* may allow an incumbent to learn faster than an equally efficient rival.

increase of keywords based on the click through rate of the first three URLs. Based on this quality measure, we find no indication for a negative average quality evolution for keywords that start from a higher initial quality level.

In this section we discuss in detail to which extent unobserved heterogeneity can confound our results. To do so, we introduce a functional form assumption that allows us to model faster learning in the *keyword dimension* through additional data in the *user dimension*. The model also allows us to take into account the impact of two potential confounding factors: (i) unobserved heterogeneity across keywords of different types and (ii) unobserved heterogeneity with respect to the age of keywords (i.e. the *time* elapsed since the keyword was first entered into the search engine). If keywords with a longer average user history are either "*easier*" or "*older*" this could explain the higher quality levels we observed in subsection 2.4.1. We show here that it is unlikely that our results are rationalized by confounding factors.

We start by introducing and discussing our functional form assumption for the quality evolution process of keywords. The quality of the search results shown for keyword $i$ at time $t$, ($Q_{it}$), is a function of the number of times the keyword has been searched till that time ($S_{it}$) and the average amount of user specific data for that keyword ($\overline{H_{it}}$):

$$Q_{it}(S_{it}, \overline{H_i}, \mu_i) = 1 - \frac{(1-\mu_i)}{S_{it}^{\delta(\overline{H_i})}} = 1 - \frac{(1-\mu_i)}{(\mathrm{T}_i \times s_i)^{\delta(\overline{H_i})}} \tag{2.5}$$

The function in Equation 2.5 is concave in $S_{it}$ and its image always lies in the interval $[0, 1]$ if $S_{it} > 1$, $\delta(\overline{H_i}) > 0$ and $\mu_i \in [0, 1]$, which we assume. $Q_{it}$ converges to 1 as $S_{it}$ approaches infinity. The function $\delta(\overline{H_i})$ determines the speed of convergences as a function of $S_{it}$. Larger values of $\delta$ imply a faster convergences to 1. More data in the *user dimension* increases the speed of learning from more searches on the same keyword if $\partial\delta(\overline{H_i})/\partial\overline{H_i} > 0$. If in turn $\partial\delta(\overline{H_i})/\partial\overline{H_i} = 0$, data on user history do not impact the speed of learning from more searches on a keyword.[16]

We denote the type of keywords by $\mu_i$, which captures the intrinsic difficulty to find relevant search results for the keyword. A larger value of $\mu_i$ corresponds to a keyword for which it is easier to retrieve relevant search results. Keywords with a larger value of $\mu_i$ start from a higher quality level and, *ceteris paribus*, remain on a higher quality level. Heterogeneity in $\mu_i$ is one source of potential confoundedness that we will study in this section. The second source of potential confoundedness stems from the fact that a search engine may have observed some keywords a longer time ago first. We will refer to $\mathrm{T}_i$ as the age of the keyword, namely the time elapsed since the search engine first ever observed that keyword. The decomposition $S_{it} = \mathrm{T}_i \times s_i$ emphasizes that the total number searches at time $t$ can written as the product of the number of months a keyword existed until time t, $\mathrm{T}_i$, and the average number of searches per month, $s_i$.[17]

Despite its simple parametric form, we believe that the function in Equation 2.5 provides a realistic approximation of statistical learning for two reasons. First, it captures diminishing returns to scale from additional searches through the concavity of the functional form in $S$. Second, the maximum achievable quality level is bounded, which captures the idea that the prediction accuracy of a model can not be increased indefinitely, i.e. that there is an irreducible error term.[18]

[16]In the functional form assumption of Equation 2.5, we neglect the possibility that $\overline{H_{it}}$ might be time varying. As discussed in Appendix 2.7.2, $\overline{H_{it}}$ should generally increases over time. Neglecting this dynamic in the functional form assumption amounts to assuming that the time varying nature of $\overline{H_{it}}$ is negligible. We consider this restriction innocuous under the detailed assumptions formulated in the data generating process in Appendix 2.7.2. It implies that differences in the age of keywords only impact quality through the resulting difference in $S_{it}$ when keywords have the same average user history in a given period.

[17]The reader should note that is corresponds to the assumption that keywords accumulate evenly over time. We refer to Appendix 2.7.2 for a detailed discussion of this assumption.

[18]See Bajari *et al.* (2019) for an extensive treatment of the properties of statistical learning.

Based on the functional form of Equation 2.5, we can analytically derive the expected quality level of keywords conditional on observing $s_i = s$ and $\overline{h_i} = h$. This mimics the scenario we face in subsection 2.4.1 and allows us to analyze under which assumptions on the distribution of $\mu$ and T we can generate data consistent with ours when the average user history does not affect the speed of learning from additional data in the *keyword dimension*.[19]

Consider the conditional expected quality, given a tuple $\{s_i = s, \overline{h_i} = h\}$:

$$\mathrm{E}[ctr_i^1|s,h] = \int_0^{\bar{\mathrm{T}}}\int_0^1 \big(1 - \frac{1-\mu_i}{(s \times \mathrm{T}_i)^{\delta(h)}}\big) f(\mu_i, \mathrm{T}_i|s,h) d\mu_i d\mathrm{T}_i \tag{2.6}$$

Where $\bar{\mathrm{T}}$ describes the maximum number of months a keyword existed before the sample period.[20] Through the conditional density function $f(\mu_i, \mathrm{T}_i|s,h)$, we can introduce correlation between the unobserved and the observed variables. We consider two scenarios. First, potential confoundedness related to the type only, which we model by assuming $f(\mu_i, \mathrm{T}_i|s,h) = f(\mu_i|s,h) \times f(\mathrm{T}_i)$. Second, potential confoundedness related to the age only, which we model by $f(\mu_i, \mathrm{T}_i|s,h) = f(\mathrm{T}_i|s,h) \times f(\mu_i)$.

We want to understand under which assumptions on $f(\mu_i|s,h)$ and $f(\mathrm{T}_i|s,h)$, the conditional expectation in Equation 2.6 generates the divergence observed in Figure 2.4 in subsection 2.4.1 if $\partial\delta(\overline{H_i})/\partial\overline{H_i} = 0$, i.e. when learning in the *keyword dimension* is not strengthened by more data in the *user dimension*. Divergence is defined in the following way: Denote by $h^l$ a long average user history and by $h^s$ a short average user history. For the difference in the conditional expectations between $h^l$ and $h^s$ given $s$, we write $\mathrm{E}[ictr_i^1|s,\Delta h] = \mathrm{E}[ictr_i^1|s,h^l] - \mathrm{E}[ictr_i^1|s,h^s]$. We observe a divergent pattern if $\mathrm{E}[ictr_i^1|s,\Delta h] > 0$ and $\partial\mathrm{E}[ictr_i^1|s,\Delta h]/\partial s > 0$, i.e. if the difference in the expected quality between $h^s$ and $h^s$ is positive and increases with $s$.

Confoundedness is modelled by first order stochastic dominance. For instance, we model the case in which a longer average user history is associated with a higher type by assuming that $F(\mu_i|s,h^s) > F(\mu_i|s,h^l)$, which implies that $\mathrm{E}(\mu_i|s,h^s) < \mathrm{E}(\mu_i|s,h^l)$. We denote the difference in the conditional expectations between $h^l$ and $h^s$ for type and age by $\mathrm{E}[\mu_i|s,\Delta h]$ and $\mathrm{E}[\mathrm{T}_i|s,\Delta h]$. Changes in the conditional expectations $\mathrm{E}[\mu_i|s,\Delta h]$ and $\mathrm{E}[\mathrm{T}_i|s,\Delta h]$ are assumed to be caused by changes in the degree of first order stochastic dominance. For instance, $\partial\mathrm{E}[\mu_i|s,\Delta h]/\partial s >= 0$ corresponds to $\partial\big[F(\mu_i|s,h^s) - F(\mu_i|s,h^l)\big]/\partial s >= 0$.

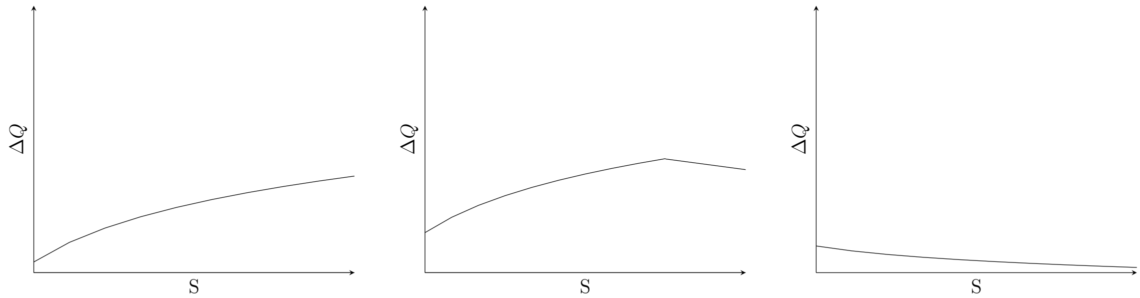
We now state our main result, which we prove in Appendix 2.7.5:

**Proposition 1.** *First, consider the case $f(\mu_i, \mathrm{T}_i|s,h) = f(\mu_i|s,h) \times f(\mathrm{T}_i)$, i.e. confoundedness in types. If $\partial\delta(\overline{H_i})/\partial\overline{H_i} = 0$, $\mathrm{E}[ictr_i^1|s,\Delta h] > 0 \quad \forall s$ and $\partial\mathrm{E}[ictr_i^1|s,\Delta h]/\partial s > 0 \quad \forall s$ can only occur if $\mathrm{E}[\mu_i|s,\Delta h] > 0 \quad \forall s$ and $\partial\mathrm{E}[\mu_i|s,\Delta h]/\partial s > 0 \quad \forall s$. Second, Consider the case $f(\mu_i, \mathrm{T}_i|s,h) = f(\mathrm{T}_i|s,h) \times f(\mu_i)$, i.e. confoundedness in age. Further assume that $\partial f(\mu_i, \mathrm{T}_i|s,h)/\partial s > 0$, i.e. more popular keywords are on average older. If $\partial\delta(\overline{H_i})/\partial\overline{H_i} = 0$, $\mathrm{E}[ictr_i^1|s,\Delta h] > 0 \quad \forall s$ and $\partial\mathrm{E}[ictr_i^1|s,\Delta h]/\partial s > 0 \quad \forall s$ can only occur if $\mathrm{E}[\mathrm{T}_i|s,\Delta h] > 0 \quad \forall s$ and $\partial\mathrm{E}[\mathrm{T}_i|s,\Delta h]/\partial s > 0 \quad \forall s$.*

Proposition 1 states that, if $\partial\delta(\overline{H_i})/\partial\overline{H_i} = 0$, a divergent pattern in $\mathrm{E}[ictr_i^1|s,\Delta h]$ can only occur if the confoundedness exactly replicates this pattern. To understand what this means, consider the case of the unobserved type of keywords. Divergence in $\mathrm{E}[ictr_i^1|s,\Delta h]$ cannot simply be explained by an average difference in type between keywords with $h = h^l$ and $h = h^s$. Instead,

[19] We focus on subsection 2.4.1 because the pattern in the data that we observe in subsection 2.4.2 can be generated under weaker assumptions on confoundedness.

[20] The reader should note that since we assume that $\overline{H_{it}}$ is not time varying, we can write $\overline{H_{it}} = \overline{H_i} = \overline{h_i} = h$.




**(a)** Average difference in quality between $h^l$ and $h^s$ if average age is small.

**(b)** Average difference in quality between $h^l$ and $h^s$ if average age is intermediate.

**(c)** Average difference in quality between $h^l$ and $h^s$ if average age is large.

**Figure 2.6:** Evolution of quality differences when data in the *user dimension* boost learning from data in the *keyword dimension*. We assume no simultaneous confoundedness.

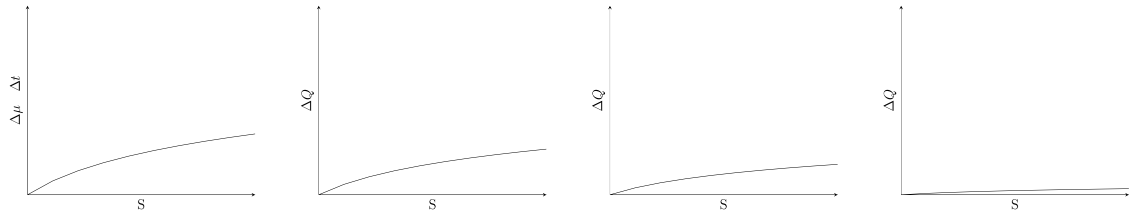
what is required is that the average difference in type between $h = h^l$ and $h = h^s$ increases with $s$. Similarly for age, any positive age difference between $h = h^l$ and $h = h^s$ would have to increase with $s$ in order to rationalize the data in the absence of network effects. Proposition 1 imposes restrictive condition on the confoundedness to generate the data we observe: The positive correlation between types or age and the average user history needs to increase monotonically with $s$.[21]

By contrast, if data obtained on the *user dimension* boost learning in the *keyword dimension*, the pattern observed in our data arises naturally. Figure 2.6 schematically describes the three main patterns we would observe in this case if the data were generated by the model introduced in this section, absent confoundedness. The pattern we observe depends on the average age of the keywords. For a young average age, we would observe a divergent pattern as the one presented in Figure 2.6a. As the average age increases, initially the pattern becomes similar to that shown in Figure 2.6b before reaching a state similar to that shown in Figure 2.6c.[22]

The main reason for the three different patterns is that all keywords eventually converge to the same quality limit. This is what explains the transition from the pattern shown in Figure 2.6a to the pattern shown in Figure 2.6b. After an initial phase of divergence in quality between $h^l$ and

[21] Note that for age confoundedness in Proposition 1, we introduce the assumption that keywords with a larger monthly search quantity need to be on average older. We consider this a weak additional restriction. An increasing age gap between keywords with a long and short average user history as a function of $s$ with a simultaneous decrease of the average age with $s$ seems artificial. Furthermore, it should also be noted that potential confoundedness between the age and average user history seems at odds with the results in subsection 2.4.2. If keywords with a longer average user history are older, on average, we would expect them to experience a smaller quality increase during our sample than keywords with a shorter average user history The same cannot be said about potential confoundedness with respect to the type, which is generally compatible with the results found in subsection 2.4.2. If keywords with a longer average user history are of a higher type, this implies a younger age than compared to keywords with a short average user history if both have the same initial quality level. This difference in age implies that keywords with a longer average user history are on the steeper part of the learning curve as compared to keywords with shorter average user history.

[22] We provide a formal discussion of the change in the pattern in Appendix 2.7.5.




**(a)** Average difference in type or age between $h^l$ and $h^s$.

**(b)** Average difference in quality between $h^l$ and $h^s$ if average age is small.

**(c)** Average difference in quality between $h^l$ and $h^s$ if average age is intermediate.

**(d)** Average difference in quality between $h^l$ and $h^s$ if average age is large.

**Figure 2.7:** Evolution of quality differences induced by unobserved heterogeneity as average age of keywords increases.

$h^s$, quality starts to converge after a certain threshold, $S^*$, is reached. Because this threshold is first reached by keywords with a large number of searches per month, we observe the pattern in Figure 2.6b for an intermediate average age of keywords. As the average age of keywords increases, the location of the kink shifts to the left, which eventually leads to a pattern similar to 2.6c. The average age that marks the transition between the different phases mainly depends on the difference in the speed of convergences between $h^l$ and $h^s$.

The functional form in Equation 2.5 implies that, when data on the *user dimension* do not impact learning from data on the *keyword dimension*, initial quality differences continuously diminish. As a consequence, we would not observe persistent differences in the data only due to unobserved heterogeneity. Observed quality differences between keywords with a long and short average user history vanish as the average age of keywords increases. Figure 2.7 illustrates how the quality differences between keywords with a long and short average user history, which are caused by confoundedness, vanish as the average age of keywords increases.[23]

The analysis of this section highlights that strong assumptions on the unobserved heterogeneity are required to generate patterns consistent with our data if the average user history does not impact the speed of learning from more searches on the same keyword. In contrast, when data in the *user dimension* boost learning in the *keyword dimension*, our model generates patterns consistent with our data if the average age of keywords is in a certain range. The results of this section suggest that simultaneous confoundedness of age and type would require assumptions similar to the one formulated for each factor of confoundedness in isolation, i.e. confoundedness between the unobservables and the average user history needs to be reinforced with $S$ in order to generate the patterns observed in the data.

## 2.6 Conclusion

In this paper we propose a mechanism that rationalizes the hypothesis of data as a source of market power. We find evidence that more comprehensive data about the users' personal search histories

[23] The pattern shown in Figure 2.7 is just one example of many potential patterns generated through heterogeneity. The main point of Figure 2.7 is to illustrate that any initial differences in quality generated through unobserved heterogeneity vanish monotonically, not to suggest a specific shape of the observed divergence.

triggers faster learning from data on searches upon the same keyword. Our results suggest that a search engine with longer user search histories is able to provide higher quality search results to the same keyword observed previously equal times.

We view the mechanism that we propose as a network effect from additional data: More data increase the efficiency of the technology, thereby causing positive externalities similar to those created by more users in a network. The search engine improves as it observes more users entering the same keyword. This is akin to the classical network effect, whereby more users improve the search result quality. In addition, we document another effect of learning from data: more data on the users' personal search histories lead to faster learning from data obtained on searches on the same keyword.

Our results provide a novel insight in the discussion on the necessary scale for a search engine to operate effectively in a competitive environment: data may provide first movers an advantage that can become difficult for rivals to overcome. This is because more data on user search histories allow the search engine to learn faster on observed click behaviour for the same keyword. Search engines with access to longer user search histories need less users to reach a given search result quality for the same keyword.

Our findings rationalize why data on users might be particularly valuable for firms operating data-driven technologies. Because data is not simply an input but also a technology shifter, our results suggest that data have the potential to confer a significant competitive advantage and to act as an barrier to entry. Our results call for awareness from antitrust policy regarding any potentially anti-competitive behaviour of firms seeking to deepening knowledge about their existing customer base. For instance, our results suggest that the benefits from locking in customers might be substantial. Similarly, merging databases across different services with a large overlap in the user base might grant firms a data advantage that is difficult to overcome for competitors. In this context, it is important to understand how market demand reacts to quality differences. This is an empirical question that remains to be answered.

Our research also reveals that there may be a strong relationship between privacy and competition policy, since the privacy interest of users may be weighed off against the externalities potentially generated through personalized information. These externalities may be positive when they enable better search results. But they may also accumulate into an advantage in data that helps a firm cement its market power. Our findings call for a careful interaction between privacy and competition policy that preserves the benefit of competition but may even allow for some data sharing to create a level playing field and tap positive externalities. In light of our findings, the right to data portability, which enables users of IT services to easily carry their personal data to other service providers, implemented in Article 20 of the EU General Data Protection Regulation, is a step in a right direction.

However, our findings also suggest that in order to substantially benefit entrants, the right of data portability would have to be exerted by a sufficient number of users switching to the new entrant. A lack of coordination and the under appreciation of the externalities involved in carrying private data to the potential entrant might result in only a few customers making use of data portability and, consequently, in sub-optimal levels of switching. A systematic theoretical and empirical assessment of user switching in light of externalities from data across users would be a valuable contribution to future research.

# 2.7 Appendix

## 2.7.1 Robustness Analysis with Alternative Quality Measures

This Appendix provides robustness checks for the results presented in sections 2.4.1 and 2.4.2 using alternative quality measures. Appendix 2.7.1.1 provides an introduction and discussion of the editorial quality measures before presenting the corresponding results in Appendix 2.7.1.2. In Appendix 2.7.1.3, we present results based on alternative click-based quality measures.

### 2.7.1.1 Editorial Quality Measures – Introduction

Our data set contains 659,000 query-URL pairs with editorial relevance judgments collected from human experts. The editorial quality judgments assess the relevance of a URL for a specific query by a categorical grade ranging from zero (*not at all relevant*) to four (*highly relevant*). By aggregating the editorial quality judgments for multiple URLs displayed on the same search result page, it is possible to obtain an overall *grade* for the quality of the result page.

Editorial quality measures are often used in the information retrieval (IR) literature to assess the quality of different algorithms in an offline environment and to compare the quality of various algorithms in an experimental setting (Chuklin *et al.*, 2013). Editorial quality measures are criticised for often struggling to capture user preferences in an online setting, which led IR researchers to develop relevance metrics based on user click behaviour (Chuklin *et al.*, 2013). We nevertheless repeat the analysis of repeat the analysis of sections 2.4.1 and 2.4.2 based on editorial quality measures.

A commonly used quality measure in the information retrieval (IR) literature is the so-called Discounted Cumulative Gain (DCG). The informational *gain* of a specific URL for a specific topic is directly assessed by the relevance grade that ranges from zero to four. The position of the URL on the result page determines by how much this informational *gain* is *discounted*. For example, assume that a specific URL is rated with a relevance grade of four relative to the searched topic. Furthermore, assume that this URL is shown on the second position of the corresponding result page. Then, we say that the *discounted gain* (DG) of this URL is given by:

$$DG = \frac{2^{rel_j} - 1}{\log_2(j+1)} = \frac{2^4 - 1}{\log_2(2+1)}, \tag{2.7}$$

Where $j$ stands for the position and $rel$ for the relevance grade of the URL. The numerator captures the informational *gain* that the user obtains by being provided this with this URL. The denominator discounts for the fact that the URL is displayed in the second position: The user had to *scan* through the search result page to be provided with this URL. Note that by applying the logarithm of base two to the denominator, the gain of a document displayed on the first position is not discounted.

To assess the quality of the entire result page, one can add up the discounted gain of all documents displayed on the first results page. Assume for convenience that all ten documents on the first result page are assigned a relevance judgment, then the discounted cumulative gain is given by:

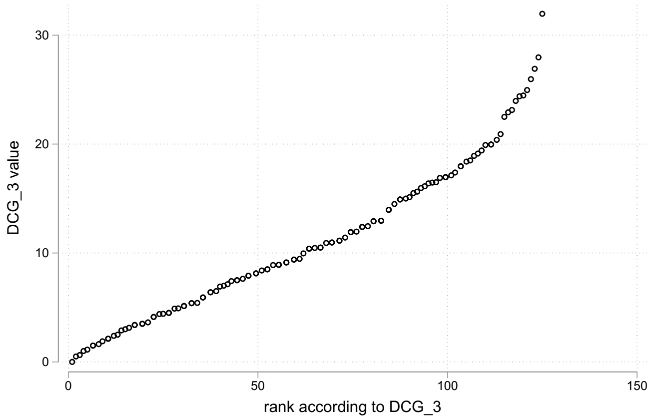



**Figure 2.8:** y-axis: $DCG^3$ value, x-axis: Ordinal ranking of relevance judgment-combinations according to their $DCG^3$ value

$$DCG^p = \sum_{j=1}^{p=10} \frac{2^{rel_j} - 1}{\log_2(j+1)}, \tag{2.8}$$

Two criteria determine the value of the DCG: (I) the general relevance of the documents available on the result page and (II) the ranking of the documents. (I) simply captures the idea that providing documents with relevant content is generally desirable (i.e. a lot of documents rated four are better than a lot of documents rated with a grade of one). (II) captures the idea that, given a specific set of documents with a given relevance, it is desirable to display the most relevant documents at the top of the result page (the ordering 4,3,2 is better than the ordering 2,3,4). The DCG captures both dimensions.

Obviously, in order to be able to compute the DCG for the entire result page, we need relevance judgments for all the URLs displayed on the page. This is only rarely the case in our dataset. Another shortcoming of the DCG measure is that the measure only allows a meaningful comparison between result pages if both pages are exactly the same number of consecutively graded URLs starting from the URL displayed on the top. It is, for example, not possible to directly compare a search result page where the first two URLs are graded with a result page where the first three URLs are graded. It is also not possible to directly compare a result page where the first three URLs are rated with a result page where only the first and third URLs are rated but the grade for the second is missing. Therefore, comparing result pages based on a DCG measure with a certain depth $p$ requires that all compared pages have URLs consecutively rated until $p$.

Consequently, the IR literature deals extensively with the imputation of relevance grades for URLs with missing grades. Usually, imputed grades are assigned based on click through rates (CTR) for the URL with the missing grade that take into consideration existing relevance grades of nearby URLs. Repeating such an exercise for the present dataset would be extremely burdensome and costly. It is for this reason that we decided to take another approach and to repeat the analysis based on two DCG-measures with different depth $p$: $DCG^1$ and $DCG^3$. The $DCG_1$ measure can be calculated for approximately 90 percent of the searches in the data set. The $DCG_3$ measure can only be computed for roughly one-third of the searches.

While the $DCG^1$ measure allows us to compute a grade for most of the searches in our data, it only takes into account the first URL. As a consequence, a results page can only be assigned 5 possible grades. While the $DCG^3$ measure allows for 125 different possible grades, it can only be computed for a fraction of the observed searches.

Each combination of relevance judgments gives rise to a particular DCG-value. These DCG-values can be used to establish an ordinal ranking of the relevance judgment combinations from 1 (lowest DCG-value) to 5 (highest DCG-value) in the case of the $DCG^1$ measure and from 1 to 125 in the case of the $DCG^3$ measure. Figure 2.8 depicts the DCG-values (y-axis) against their ordinal ranking (x-axis) for the $DCG^3$ measure.

Figure 2.8 illustrates the convex nature of the DCG-measure: the difference in DCG-values between the relevance judgment combination (4,4,4) and (4,4,3) is larger than the difference in DCG-values between the relevance judgment combination (0,0,1) and (0,0,0). In other words, incremental improvements of relevance judgment combinations lead to higher DCG increases as we move along the ordinal ranking of URL-combinations. This is due to the fact that the relevance judgments enter the DCG formula in the exponent.

This property is mechanical rather than informative about the true added quality gain for customers. In the above example it is debatable, whether the improvement from (0,0,0) to (0,0,1) is more or less valuable to the consumer than the improvement from (4,4,3) to (4,4,4). Classical economic thinking would suggest that the former is more valuable than the latter, if relevance judgments could be interpreted directly as “information-units”.

If we use the ordinal ranking of the URL-combinations each incremental improvement is valued the same. An improvement for a given keyword would then be larger, the more “steps” it improved on the ordinal ranking. This choice seems the most sensible to us. It is for this reason that we opt for the ordinal rank dictated by the DCG measure rather than for the DCG measure itself to perform our analysis. For the remainder of the analysis, we refer to the editorial quality measures as $rank^1$ and $rank^3$, respectively. The correlation coefficient between $rank^1$ and $ctr^1$ is 0.55. The correlation coefficient between $rank^3$ and $ctr^1$ is 0.49.

#### 2.7.1.2 Editorial Quality Measures – Results

Figure 2.9 shows the results of estimating Equation 2.3 on the grid $s \times h$ with $rank^1$ as the dependent variable. In Figure 2.9a, estimation is performed for all keywords in the sample. In Figure 2.9b estimation is performed for the subset of keywords that do not deviate by more than ten percentage point from the even accumulation criterion. Figure 2.10 shows the corresponding analysis with $rank^3$ as the dependent variable.

For both editorial quality measures, only minor changes to the observed pattern are caused by dropping keywords that see less steady search patterns over time. The changes are larger when we perform the same exercise using the click-based quality measures and removal keywords that see more rapid boosts in searches on some days.

Nevertheless, Figures 2.11 and 2.12 reveal that the effect of removing keywords do not accumulate evenly increases the measured impact of the network effect for both editorial quality measures in a similar way as does the click-based quality measure. Furthermore, the divergent pattern in the quality levels between keywords with a long and short average user history that we observe for the click-based quality measures is also present when using the editorial quality measures.

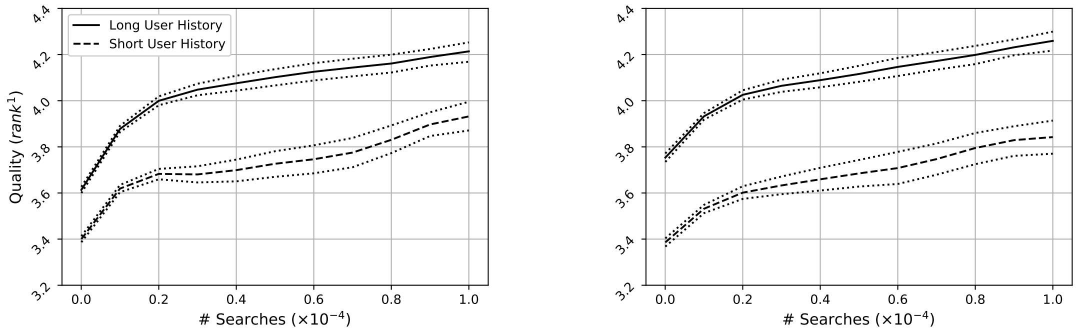
Figures 2.13 and 2.14 show the results of estimating Equation 2.4 on the grid $ictr \times h \times s$ for $rank^1$ and $rank^3$, respectively. For both editorial quality measures, we normalized the quality to lie in the interval $[0, 1]$ when analyzing the quality evolution of keywords.

In both Figure 2.13 and Figure 2.14, we observe that the measured quality increase generally declines with the initial quality level. Furthermore, the quality increase of keywords with a long average user history is always weakly larger than the quality increase measured for keywords with a short average user history.

For the $rank^1$ measure, we observe no difference between long and short keywords for the lowest initial quality level. The large confidence intervals reflect the fact that only few keywords start from this quality level in our sample. For the second highest initial quality level, the observed pattern roughly corresponds to our expectations of differential learning speeds between keywords with a long and short average user history. A general problem that we see with the $rank^1$ measure is its very crude scale, which only allows for five different quality grades. If differential learning between keywords with a long and short average cookie length occurs on a more granular level within the sample period, the $rank^1$ measure is likely to poorly capture differential learning.

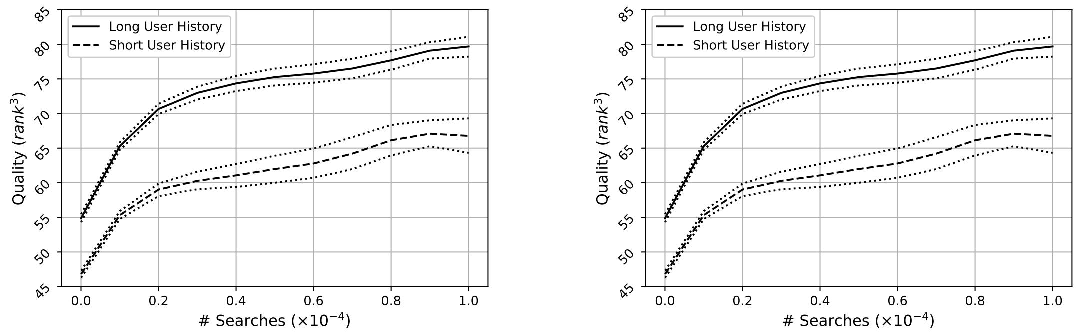
For the $rank^3$ measure, the difference in learning between keywords with a long and short average user history is most pronounced for the lowest initial quality level. For larger initial quality levels, the difference is not pronounced. The large confidence intervals that we observe for all initial levels reflect the loss of observations associated with using the $rank^3$ quality measure.

Altogether, the results obtained with the editorial quality measures are in line with the results obtained using the click-based quality measures. The divergence of the quality levels observed in Figures 2.11 and 2.12, and the differential learning speed observed in Figures 2.13 and 2.14 for lower initial quality levels are in line with the hypothesis that additional data in the *user dimension* boosts learning from data in the *keyword dimension*.




**(a)** Entire sample of keywords **(b)** Keywords accumulating evenly

**Figure 2.9:** Average quality of keywords with 95 % confidence intervals




**(a)** Entire sample of keywords **(b)** Keywords accumulating evenly

**Figure 2.10:** Average quality of keywords with 95 % confidence intervals

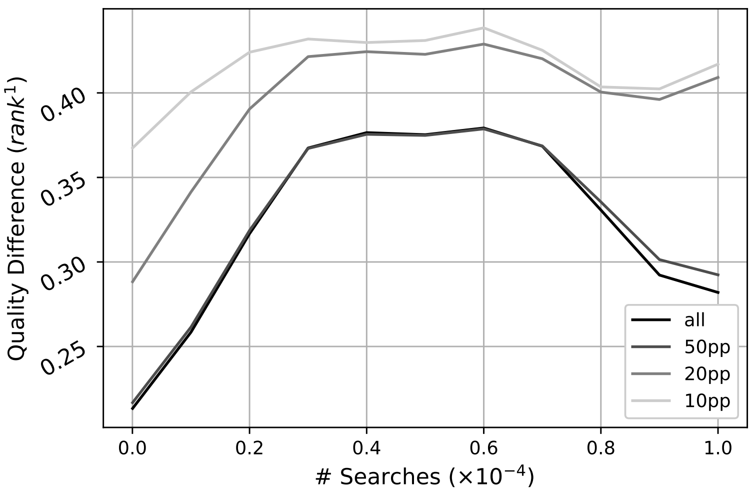



**Figure 2.11:** Impact of longer average user history on quality level. Different tolerance levels for deviation from even accumulation criterion are considered.

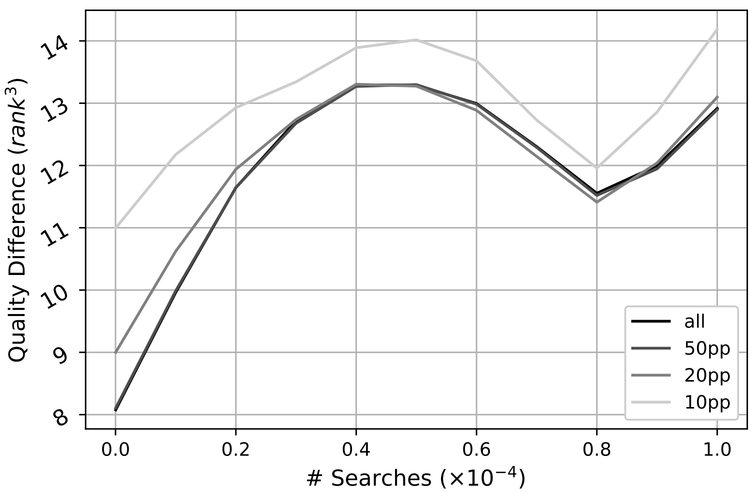



**Figure 2.12:** Impact of longer average user history on quality level. Different tolerance levels for deviation from even accumulation criterion are considered.

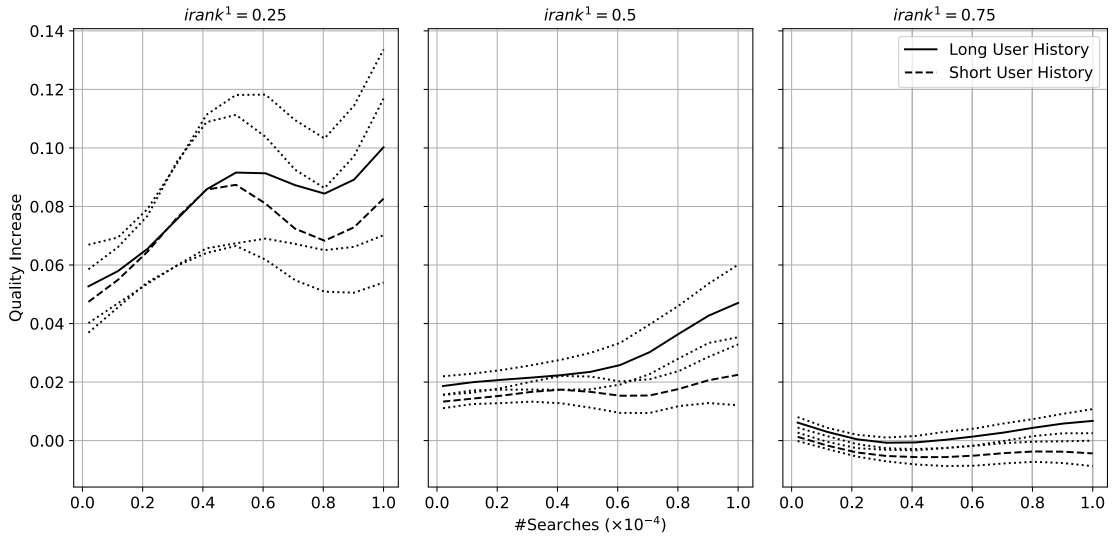



**Figure 2.13:** Average quality increase of keywords with 95% confidence intervals. Each panel stands for a different initial quality level.

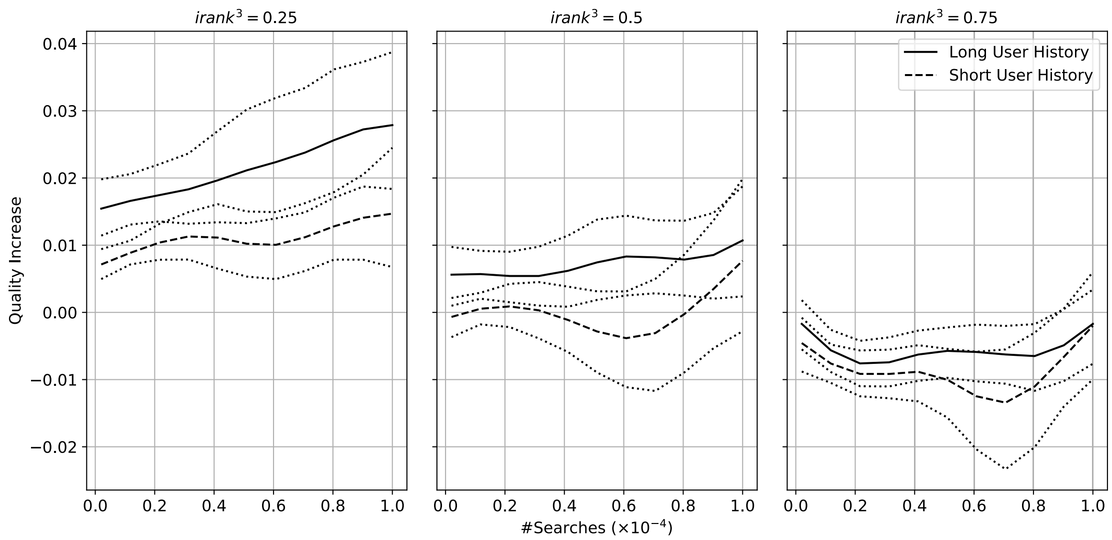



**Figure 2.14:** Average quality increase of keywords with 95% confidence intervals. Each panel stands for a different initial quality level.

#### 2.7.1.3 Alternative Click Based Quality Measures – Results

In this Appendix, we present the results obtained with two alternative click-based quality measures. For the first, search result quality is encoded as "good" if the last recorded click occurs on one of the first three organic URLs. For the second, search result quality is encoded as "good" if the last recorded click occurs on one of the first ten organic URLs. We denote the first alternative click based measure by $ctr^3$ and the second by $ctr^{all}$. For $ctr^{all}$, search result is only considered " bad" if the searcher ends the search on the second result page or leaves the search engine without finding a URL she considers relevant.

**Table 2.1:** Correlation click based vs. editorial measures

| | $ctr^1$ | $ctr^3$ | $ctr^{all}$ |
|---|---|---|---|
| $rank^1$ | 0.55 | 0.51 | 0.41 |
| $rank^3$ | 0.49 | 0.45 | 0.34 |

From Table 2.1, it can be seen that the correlation between the click based quality measures and each editorial quality measure decreases as the criterion for a "good" search result quality is defined more broadly. Furthermore, each click based quality measure correlates more highly with the editorial quality measure of depth one.

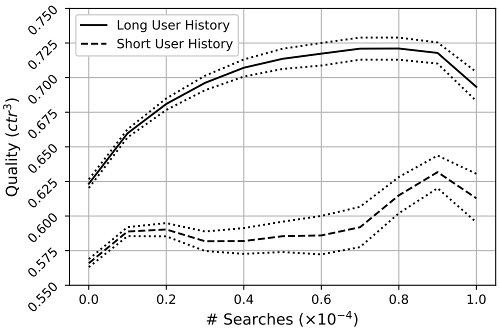
Figure 2.15 shows the results of estimating Equation 2.3 on the grid $s \times h$ with $ctr^3$ as the dependent variable. The results in Figure 2.15a are based on all the keywords in the sample. Figure 2.15b shows the results for the population of keywords that deviate no more than ten percentage points from the even accumulation criterion. Figure 2.16 repeats the same exercise for the $ctr^{all}$ quality measure. Figures 2.17 and 2.18 shows the consequence of continuously narrowing the tolerance interval for the even accumulation criterion on the measured network effects for the $ctr^3$ and $ctr^{all}$ quality measure, respectively. All Results are qualitatively identical to the results obtained based on the $ctr^1$ quality measure.

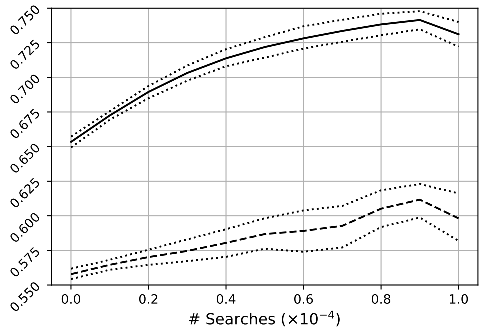
Figures 2.19 and 2.20 show the results of estimating Equation 2.4 on the grid $ictr \times h \times s$ for $ctr^3$ and $ctr^{all}$, respectively. The results are qualitatively identical to the ones obtained based on the $ctr^1$ quality measure.

Its is noteworthy that, in general, the measured network effect between keywords with a long and short average user history tends to decrease as we broaden the set of URLs that determine a "good" search result quality. We find this result intuitive: It requires less personalized knowledge to get three out of ten or ten out of ten results right as it requires to get one out of ten results. A more broadly defined quality measure therefore de-emphasizes the value of personalized information.

Another noteworthy difference to the results in the main section is that the quality evolution during the sample period is now also positive for keywords with a high initial quality level and a short average user history. This indicates that the decrease in the click through rate on the first URL, which we found in Section 2.4.2, corresponds to an increase in the click through rate for URLs further down the results list, i.e. that URLs further down the result list cannibalize clicks on the first URL. Again, it seems intuitive to see this phenomenon occur for keywords with a short average user history, where the search engine lacks the information to target individual preferences.

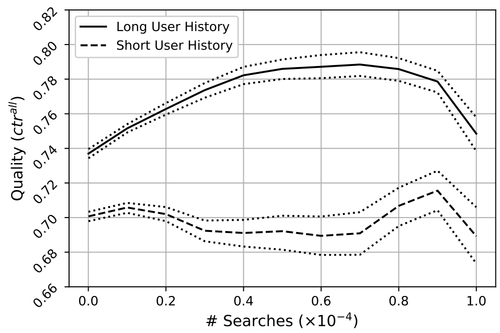
Altogether, the results based on the alternative click-based quality measure prove to be robust to the results presented in the main section of the paper. Minor differences to the results in the main section appear to be sensible.

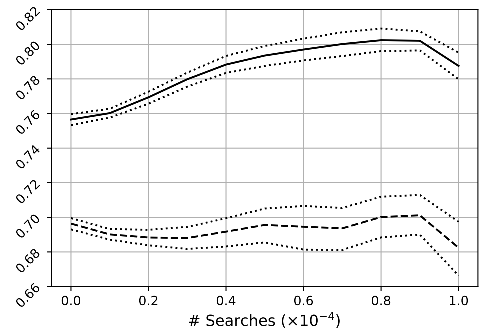



**(a)** Entire sample of keywords **(b)** Keywords accumulating evenly

**Figure 2.15:** Average quality of keywords with 95 % confidence intervals




**(a)** Entire sample of keywords **(b)** Keywords accumulating evenly

**Figure 2.16:** Average quality of keywords with 95 % confidence intervals

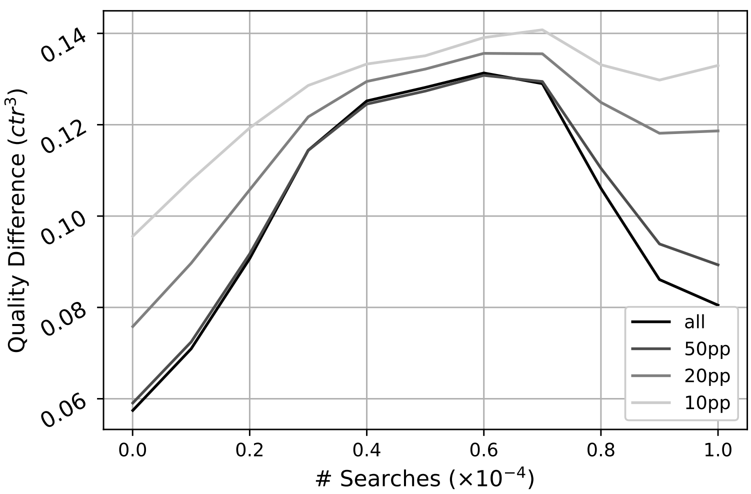



**Figure 2.17:** Impact of longer average user history on quality level. Different tolerance levels for deviation from even accumulation criterion are considered.

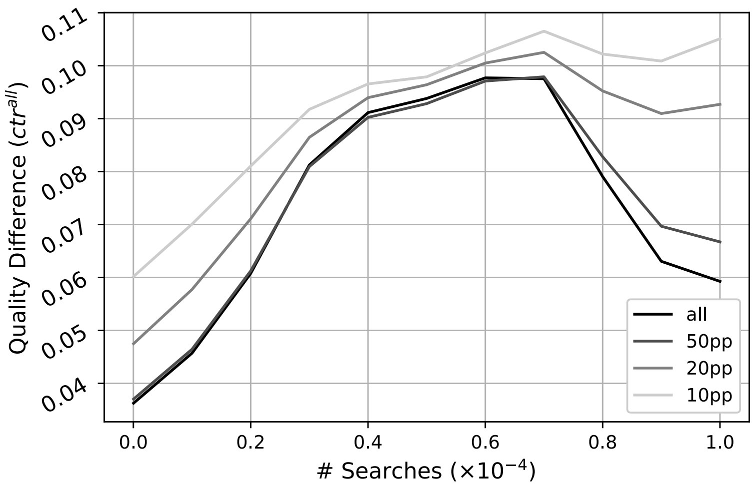



**Figure 2.18:** Impact of longer average user history on quality level. Different tolerance levels for deviation from even accumulation criterion are considered.

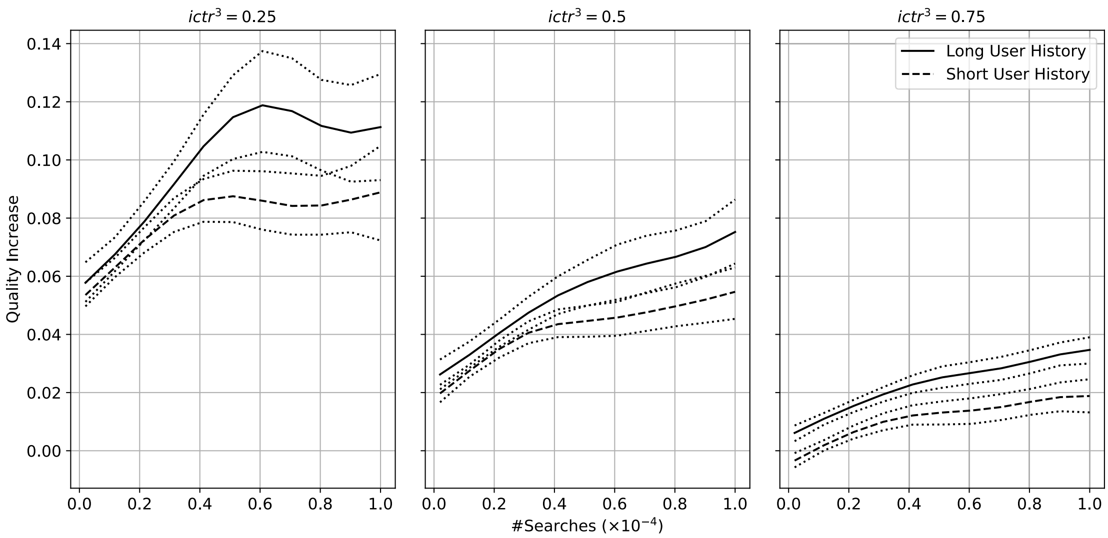



**Figure 2.19:** Average quality increase of keywords with 95% confidence intervals. Each panel stands for a different initial quality level.

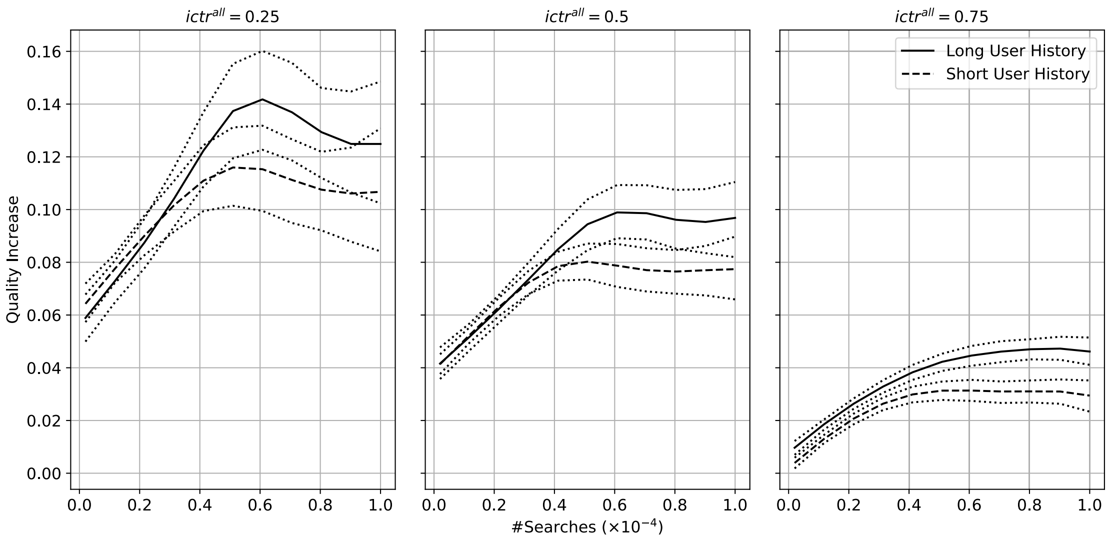


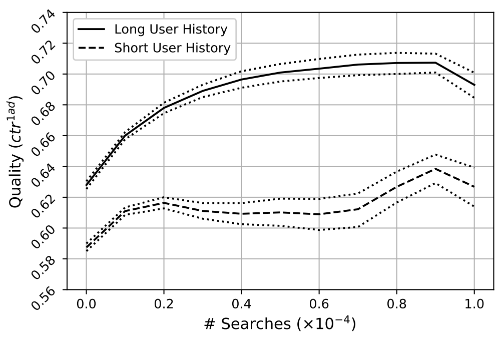

**Figure 2.20:** Average quality increase of keywords with 95% confidence intervals. Each panel stands for a different initial quality level.

#### 2.7.1.4 Counting clicks on advertisement URLs – Results

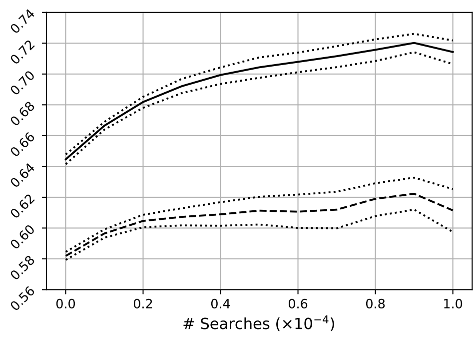
This Appendix gives the results for the robustness checks when search result quality is encoded as *good* if the last click occurs on the first URL or an Advertisement URL. The click through rate is now computed as follows:

$$ctr_i^{\{1ad\}} = \frac{\sum_{s_i \in S_i} \mathbb{1}\{lcp_{is} \in \{0,1\}\}}{\sum_{s_i \in S_i} \mathbb{1}\{lcp_{is} = \Omega\}} \tag{2.9}$$

In comparison to the other click based quality measures used in the paper, the denominator now sums up all the searches in $S$. The numerator now considers clicks on advertisement URLs as "good" search result quality. For completeness, it should be noted that the data description offers no clear guidance on the exact nature of the clicks that we characterize as "ads". These clicks could be clicks on ads but also clicks on spelling suggestions or reformulations of the original keyword the user submitted. The results in Figures 2.21, 2.22 and 2.23 show that our conclusions remain unaffected by whether or not, we count these clicks or not.




**(a)** Entire sample of keywords




**(b)** Keywords accumulating evenly

**Figure 2.21:** Average quality of keywords with 95 % confidence intervals

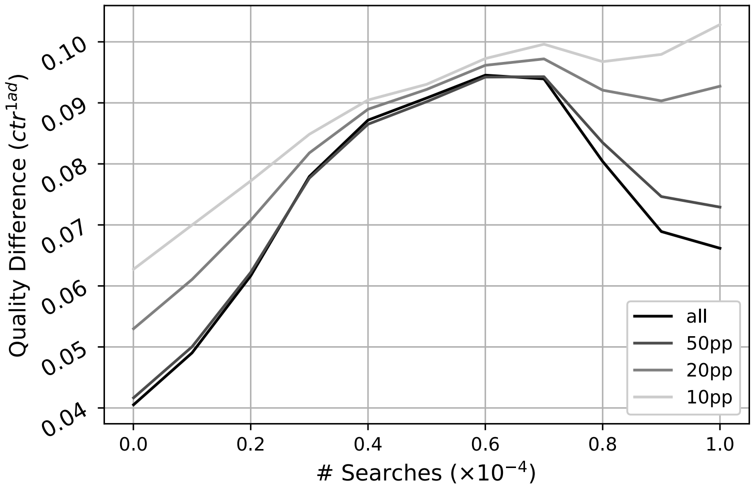



**Figure 2.22:** Impact of longer average user history on quality level. Different tolerance levels for deviation from even accumulation criterion are considered.

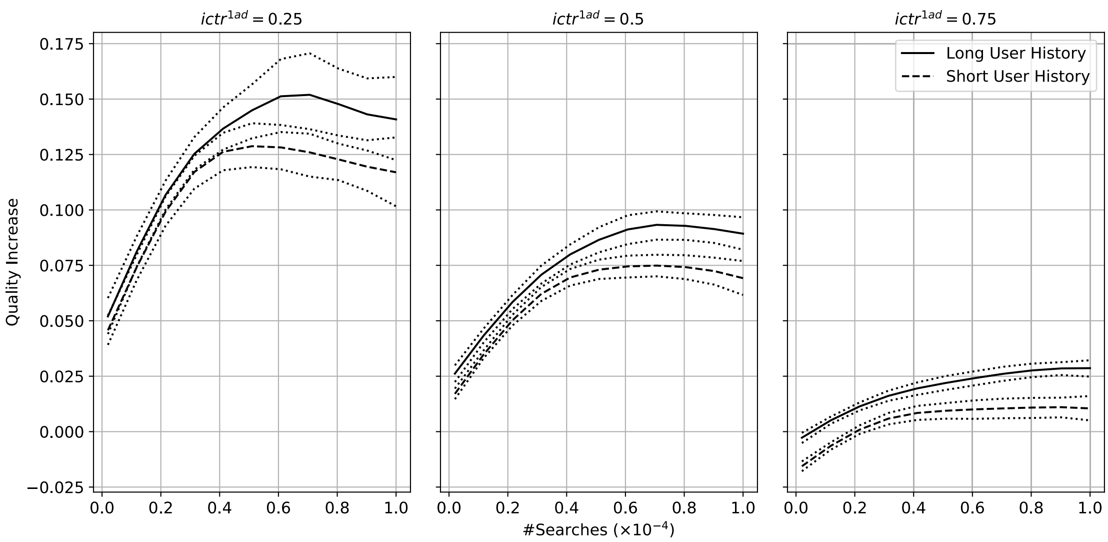



**Figure 2.23:** Average quality increase of keywords with 95% confidence intervals. Each panel stands for a different initial quality level.

### 2.7.2 Formal Discussion of Data Generating Process

In subsection 2.4.1, we use the variables we observe, $s_i$ and $\overline{h_i}$, as proxy measures for $S_{it}$ and $\overline{H_{it}}$ and demonstrate that keywords for which the proxies indicate larger values of $S_{it}$ and $\overline{H_{it}}$ are on a higher quality level. In this Appendix, we are interested in the formal relationship between $s_i$ and $\overline{h_i}$ and their unobserved counterparts at time $\underline{t}$, which denotes the beginning of our sample period.

Reducing the approximation error we make from substituting $S_{i\underline{t}}$ and $\overline{H_{i\underline{t}}}$ by their proxy variables $s_i$ and $\overline{h_i}$ amounts to reducing $\text{Var}[S_{i\underline{t}}|s, h]$ and $\text{Var}[\overline{H_{i\underline{t}}}|s, h]$ for each tuple $\{s_i = s, \overline{h_i} = h\}$ that we condition on. Conditioning on a tuple $\{s_i = s, \overline{h_i} = h\}$ induces a distribution over $S_{i\underline{t}}$ and $\overline{H_{i\underline{t}}}$. The expectations of these distributions, $\text{E}[S_{i\underline{t}}|s, h]$ and $\text{E}[\overline{H_{i\underline{t}}}|s, h]$, describe the expected values of $S_{i\underline{t}}$ and $\overline{H_{i\underline{t}}}$ for each tuple $\{s_i = s, \overline{h_i} = h\}$. The variances describe the heterogeneity over $S_{i\underline{t}}$ and $\overline{H_{i\underline{t}}}$ for each tuple $\{s_i = s, \overline{h_i} = h\}$. The larger the variances, the larger the error we make from substituting $S_{i\underline{t}}$ and $\overline{H_{i\underline{t}}}$ by $s_i$ and $\overline{h_i}$.

We now introduce a generating process under which $\text{Var}[S_{i\underline{t}}|s, h]$ and $\text{Var}[\overline{H_{i\underline{t}}}|s, h]$ are minimized.

**DGP 1:** Keywords originate at random time and have constant monthly popularity. i.e. the number of searches for the keywords is constant in every month. I.e. $s_{i\tau} = s_i \quad \forall \tau$, where $\tau$ denotes a month.

**DGP 2:** If two keywords follow **DGP 1** and have the same average user history in month $\tau$, they also have the same average user history in month $\tau + 1$.

To understand why **DGP 1** and **DGP 2** reduce the variance, it is crucial to note that there are two sources of heterogeneity that (potentially) influence $\text{Var}[S_{i\underline{t}}|s, h]$ and $\text{Var}[\overline{H_{i\underline{t}}}|s, h]$: (i) the number of months a keyword existed before the period of our sample, and, (ii) the monthly variability of $s_{i\tau}$ (for $\text{Var}[S_{i\underline{t}}|s, h]$) and $\overline{h_{i\tau}}$ (for $\text{Var}[\overline{H_{i\underline{t}}}|s, h]$). The conditions formulated in **DGP 1** and **DGP 2** eliminate the second source of heterogeneity.

By **DGP 1**, all keywords for which we observe a given number of searches during the sample period had the same monthly popularity in all previous periods. By **DGP 2**, all keyword for which we observe a given average user history during our sample also had the same average user history in all previous periods. Therefore, if we can determine a population of keywords for which **DGP 1** and **DGP 2** is a plausible approximation, we can approximate $S_{i\underline{t}}$ and $\overline{H_{i\underline{t}}}$ more accurately through $s_i$ and $h_i$.[24]

**DGP 2** can be understood in the following way: keywords differ in the type of user who search the keyword. One dimension of the user type is the usage intensity of the search engine, which determines $\overline{H_{it}}$. **DGP 2** states that if two keywords have the same average user type in one month, they will also have the same average user type in the next month. Since neither experience a change in popularity (note that we assume that **DGP 1** holds), this is a sensible assumption to make since no change in popularity indicates that the user type did not change. By contrast if **DGP 1** does not hold, it is also likely that the user type changed and therefore that **DGP 2** does not hold. For the sake of this analysis, we assume that if **DGP 1** is violated than **DGP 2** cannot hold.

Note that **DGP 2** states a weaker assumption than **DGP 1** because it does not require that the average user history is constant. While a constant monthly popularity seems a realistic approximation for the evolution of the total search quantity of a keyword that reached its steady state,

[24]It should be noted that, to simplify analysis, we abstract from the fact that $s_i$ and $\overline{h_i}$ are subject to sampling error.

a similar assumption for the average user history appears unrealistic. This is because the search engine continuously collects data on users, which suggest that the average user history should have tendency to increase.

Given the above reasoning, it is sufficient to develop a method to detect keywords for which **DGP 1** is a poor approximation. By removing these keywords, we can focus on a population of keywords for which $\text{Var}[S_{i\underline{t}}|s,h]$ and $\text{Var}[\overline{H_{i\underline{t}}}|s,h]$ is minimal. By **DGP 1**, we need to drop keywords for which the monthly popularity that we observe is a unreliable measure for the monthly popularity in previous periods. Because we have no information on the previous periods, we rely on the simple heuristic that keywords that experience popularity changes during our sample period do likely not follow **DGP 1**.

The heuristic relies entirely on the intuition that popularity changes in the sample are indicative of long term patterns in popularity: A keyword that increases in popularity during each day in our sample is likely on a long run upward trend. Similarly, a keyword on a downward trend in our sample is likely also on a long run downward trend. Both patterns, upward or downward trends in our sample, could also be indicative of an oscillating long run popularity, with phases of high and low popularity. The key point is that all mentioned scenarios are inconsistent with **DGP 1**.

Next, we provide formal arguments why **DGP 1** and **DGP 2** reduce the variances. We discuss each assumption separately.

#### 2.7.2.1 DGP 1

Throughout, we assume that keywords originate at random time in the past. Denote by $\mathrm{T}_i$, the number of months a keywords existed previous to the sample period. We denote the average monthly popularity of a keyword by $s_i$. Note that, by construction, we have $S_{i\underline{t}} = \mathrm{T}_i \times s_i$. The total number of searches previous to the month our data were sampled is equal to the number of months a keywords existed previous to the sample period times the average monthly popularity.

Therefore, for the population of keywords that do not follow **DGP 1**, we have that $\text{Var}[S_{i\underline{t}}|s,h] = \text{Var}[s_i \times \mathrm{T}_i|s,h]$. If we make the assumption of conditional independence between $s_i$ and $\mathrm{T}_i$, the conditional variance of the product $s_i \times \mathrm{T}_i$ is given by $\sigma_s^2\sigma_{\mathrm{T}}^2 + \sigma_s^2\mu_{\mathrm{T}}^2 + \mu_s^2\sigma_{\mathrm{T}}^2$. Where $\sigma$ denote conditional variances and $\mu$ conditional expectations. For a population of keywords that follows **DGP 1**, we have $\sigma_s^2 = 0$ and, consequently, the variance reduces to $\mu_s^2\sigma_{\mathrm{T}}^2$, which is obviously smaller.

The assumption that we must make above, in addition to conditional independence, is that $\mu_s$ and $\sigma_{\mathrm{T}}^2$ do not change when dropping keywords that do not follow **DGP 1**. This amounts to assuming that keywords that do not follow **DGP 1** have $s_i$ that randomly deviate from $\mu_s$. Furthermore, they must be drawn from the same distribution of $\mathrm{T}_i$.

#### 2.7.2.2 DGP 2

For **DGP 2**, first note that we can write $\overline{H_{i\underline{t}}} = \sum_1^{\mathrm{T}_i} w_\tau \times h_\tau$, where $\tau$ denote months previous to the sample period and $h_\tau$ is the average user history in month $\tau$. $w_\tau$ is the monthly weight given by $s_\tau / S_{i\underline{t}}$, where $s_\tau$ is the number of searches in month $\tau$. We have $\sum_1^{T_i} w_\tau = 1$. Thus, the average user history previous to the sample period is the weighted monthly average user history, where the weights are the ratio of the number of searches in a given month in the total number of searches previous to sample period.

If we would assume that $h_\tau = h \quad \forall \tau$, then $\overline{H_{i\underline{t}}} = h$ and by **DGP 2**, $\text{Var}[\overline{H_{i\underline{t}}}|s,h] = 0$. While we find it realistic to assume that the monthly popularity of keywords $s_\tau$ stays roughly constant, as stated in **DGP 1**, we do not believe it is a realistic assumption for the average user history. Instead, it is more realistic to assume that the average user history increases over time as users continuously interact with the search engine. This is why **DGP 2** is formulated more broadly.

**DGP 2** states that for a given $h$ there is a unique sequence of monthly average user histories previous to the sample period: $\{h_1, ..., h_\tau, h_{\tau+1}, ..., h_{\mathrm{T}_i}\}$. The only source of heterogeneity between keywords is the length of the sequence determined by $\mathrm{T}_i$.

Thus, if **DGP 2** holds, for two keywords $i$ and $j$ with $\mathrm{T}_i \neq \mathrm{T}_j$, we have that $h_{\tau i} = h_{\tau j}$ for all $\tau \leq min\{\mathrm{T}_i, \mathrm{T}_{ij}\}$. If **DGP 2** would not hold, in addition to the heterogeneity in the length of the sequences, we would in general have that $h_{\tau i} \neq h_{\tau j}$ for $\tau \leq min\{\mathrm{T}_i, \mathrm{T}_{ij}\}$. It is logical that this additional heterogeneity can only increase $\mathrm{Var}[\overline{H_{i\underline{t}}}|s, h] = 0$.

### 2.7.3 Details on Kernel Estimation

Throughout the paper, estimation is performed by means of local linear regression. The employed kernel for the weighting of the observations is a radial Gaussian kernel. The bandwidth of the kernel determines the standard deviation of the Gaussian distribution used for weighting, which is identical for each dimension. The off-diagonal elements of the covariance matrix are set to zero. For estimation, all explanatory variables are normalized to lie in the interval between zero and one. Because all explanatory variables only take positive values, this transformation amounts to dividing each explanatory variable by its maximum value.

For the analysis of subsection 2.4.1 and all the associated robustness checks in appendices 2.7.1.2 and 2.7.1.3, the bandwidth is set to 0.1, irrespective of the employed quality measure and the considered subsample resulting from dropping keywords that do not follow **DGP 1** and **DGP 2**. For the analysis of Section 2.4.2 and all associated robustness checks, in appendices 2.7.1.2 and 2.7.1.3, the bandwidth is set to 0.15.

Table 2.2 reports the optimal bandwidths determined by leave one out cross validation for each quality measure/subsample combination considered in the analysis of subsection 2.4.1 and the corresponding robustness checks in appendices 2.7.1.2 and 2.7.1.3. Table 2.3 reports the optimal bandwidths determined by leave one out cross validation for each quality measure for the analysis in subsection 2.4.2 and the corresponding robustness checks in appendices 2.7.1.2 and 2.7.1.3.

Throughout the analysis, we trade reduced variance for increased bias, which means that we selected a bandwidth larger than the optimal bandwidth determined by leave one out cross validation. As shown in Tables 2.2 and 2.3, the bandwidth chosen for the analysis in subsection 2.4.1 is further away from the optimal bandwidth determined by cross validation than the bandwidth we chose for the analysis in subsection 2.4.2.

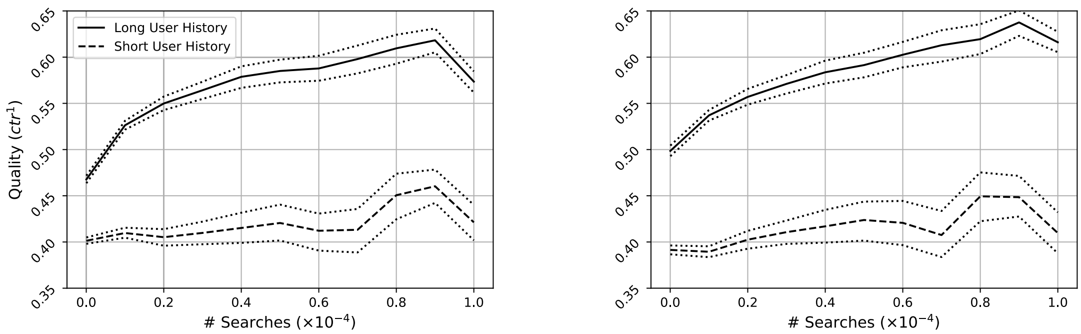
Figures 2.24 and 2.25 show the results for the analysis performed in subsection 2.4.1 when employing a bandwidth of 0.05, which is close the optimal bandwidth determined by the cross validation procedure. The irregular patterns observed in Figures 2.24a and 2.24b indicate that the selected optimal bandwidth tends to over-fit the data. Figures 2.24 and 2.25 also demonstrate that the results in subsection 2.4.1 are not driven by the selection of a wider bandwidth, which is reassuring.

**Table 2.2:** Optimal bandwidths for analysis of section 2.4.1

| | all | dev:50pp | dev:20pp | dev:10pp |
|---|---|---|---|---|
| $ctr^1$ | 0.03 | 0.03 | 0.04 | 0.04 |
| $ctr^3$ | 0.03 | 0.03 | 0.04 | 0.05 |
| $ctr^{all}$ | 0.03 | 0.03 | 0.04 | 0.05 |
| $rank^1$ | 0.04 | 0.04 | 0.06 | 0.1 |
| $rank^2$ | 0.04 | 0.04 | 0.04 | 0.05 |
| $rank^3$ | 0.04 | 0.06 | 0.1 | 0.08 |

**Table 2.3:** Optimal bandwidths for analysis of section 2.4.2

| | dev:50 pp |
|---|---|
| $ctr^1$ | 0.11 |
| $ctr^3$ | 0.10 |
| $ctr^{all}$ | 0.09 |
| $rank^1$ | 0.06 |
| $rank^2$ | 0.10 |
| $rank^3$ | 0.14 |




**(a)** Entire sample of keywords **(b)** Keywords accumulating evenly

**Figure 2.24:** Average quality of keywords with 95% confidence intervals

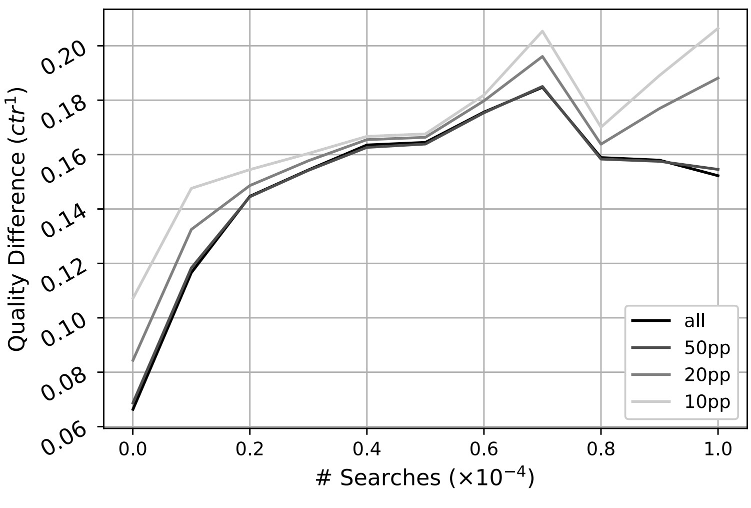



**Figure 2.25:** Impact of longer average user history on quality level. Different tolerance levels for deviation from even accumulation criterion are considered.

### 2.7.4 Regression to the Mean (RTM)

The regression to the mean (RTM) problem arises in any analysis in which observations are classified based on an initial outcome (i.e based on the dependent variable). Intuitively, the problem arises because subjects are "erroneously" allocated to a category based on a single (or few) observation(s), which is not representative for the average of that subject. Because over time the outcome associated with a subject tend to revert to the average value commonly observed for that subject, studies that rely on a classification based on an initial outcome are prone to be affected by the RTM effect.

In the context of the analysis in subsection 2.4.2, there is the concern that the initial quality of keywords is a bad measure for the true initial quality of the keywords. If a keyword is assigned to the low initial quality group "by chance", because it happened to experience a large idiosyncratic deviation from its true quality in the beginning of the sample period, it will naturally revert to its true average quality subsequently. This might lead to the erroneous conclusion that keywords with a low initial quality display a high quality increase, whereas in reality they simply revert to their natural average. Intuitively, the problem is more pronounced if we rely on a few observations to assess the quality; i.e. if we rely on an imprecise measurement of the true initial quality.

Figure 2.26 shows the result of the analysis of subsection 2.4.2 when we compute the initial quality of the keywords based on the searches a keyword experiences during the first day of the sample. In one day, a keyword with 200 searches experiences approximately six searches on average (200/32). Note that in subsection 2.4.2, we use the first 100 searches to determine the initial quality. As shown in Figure 2.26, the estimated quality increase is especially " irregular" for keywords with a low total search quantity: Keywords with a low initial quality experience a quality increase of roughly 10 percentage points.

To assess to what degree the results are driven by regression to the mean, we apply the correction formula discussed in Barnett *et al.* (2004). This formula is derived under the assumption of normally distributed stationary data and when observations are classified based on thresholds. For example, the formula can be applied in treatment analysis when interested in the effect for individuals who are below a certain income threshold. The underlying assumption is that, absent treatment, there is no significant income trend and that income is normally distributed for the studied population. To the best of our knowledge, no formula exists for our specific application.

Our analysis does not rely on thresholding and our data are highly non-normally distributed and non-stationary. Our kernel estimation approach defines weights for observations close to the kernel centroid. Observations are assigned less weight the further away they are from the kernel centroid. To apply the formula, we ignore the weighting and estimate the average RTM effect as if all observations are equally weighted. However, we only include observations for which the average user history and the total number of searches are within one standard deviation of the kernel centroid.

More precisely, imagine we want to approximate the RTM effect for the estimate at the kernel centroid defined by an initial quality level of $ctr^1 = 25\%$ $s = 2000$ and $h = Q_1$. Imagine that in the estimation we specified a standard deviation of 0.1 for the radial Gaussian kernel. To approximate the RTM, we apply the formula in **?** to all observations for which the Euclidean distance to the kernel centroid in the space spanned by $s$ and $h$ is less than or equal to 0.1. By doing so, the RTM is calculated for the observations that had most weight during the estimation.

The left column of Figure 2.27 displays the approximated RTM effect for the results presented in Figure 2.26. As shown, the RTM effect seems to explain much of the observed quality changes observed for small search quantities in Figure 2.26. The right column of Figure 2.27 displays the approximated RTM for the analysis presented in subsection 2.4.2. As can be seen, the regression to the mean effect appears to be much less pronounced when we use 100 observations to estimate the initial quality.

At around $4,000$ searches the estimated RTM effect in the left column of 2.27 has roughly the same magnitude as the RTM effect in the right column of Figure 2.27. This is the range in which the number of searches in one day corresponds to roughly 100 searches. Our heuristic to assess the RTM effect seems to deliver sensible results. When both methods to assess the initial quality rely on the same number of searches, the estimated RTM effect is similar. It is also worthwhile to note that the estimated effect seems mostly only marginally different between keywords with a long and short average user history . This lessens the concern that the observed differences in the quality increase between keywords with a long and short average user history observed subsection 2.4.2 are driven by a differential impact of the RTM effect.

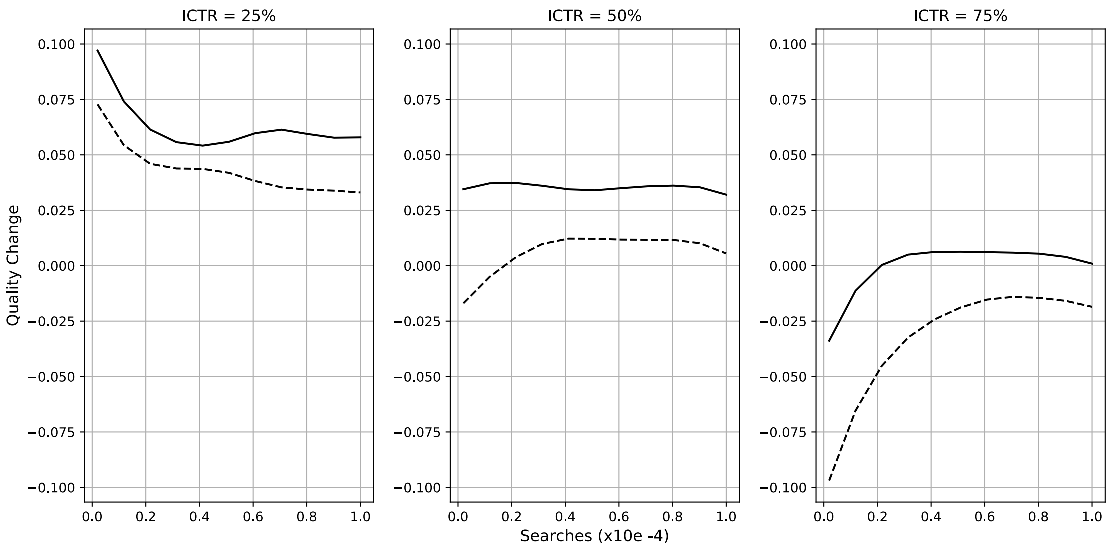



**Figure 2.26:** Analogue results to Figure 2.5 of subsection 2.4.2 if analysis is performed based on quality differences between first and last day and the initial quality is assessed based on the click through rate on the first day.

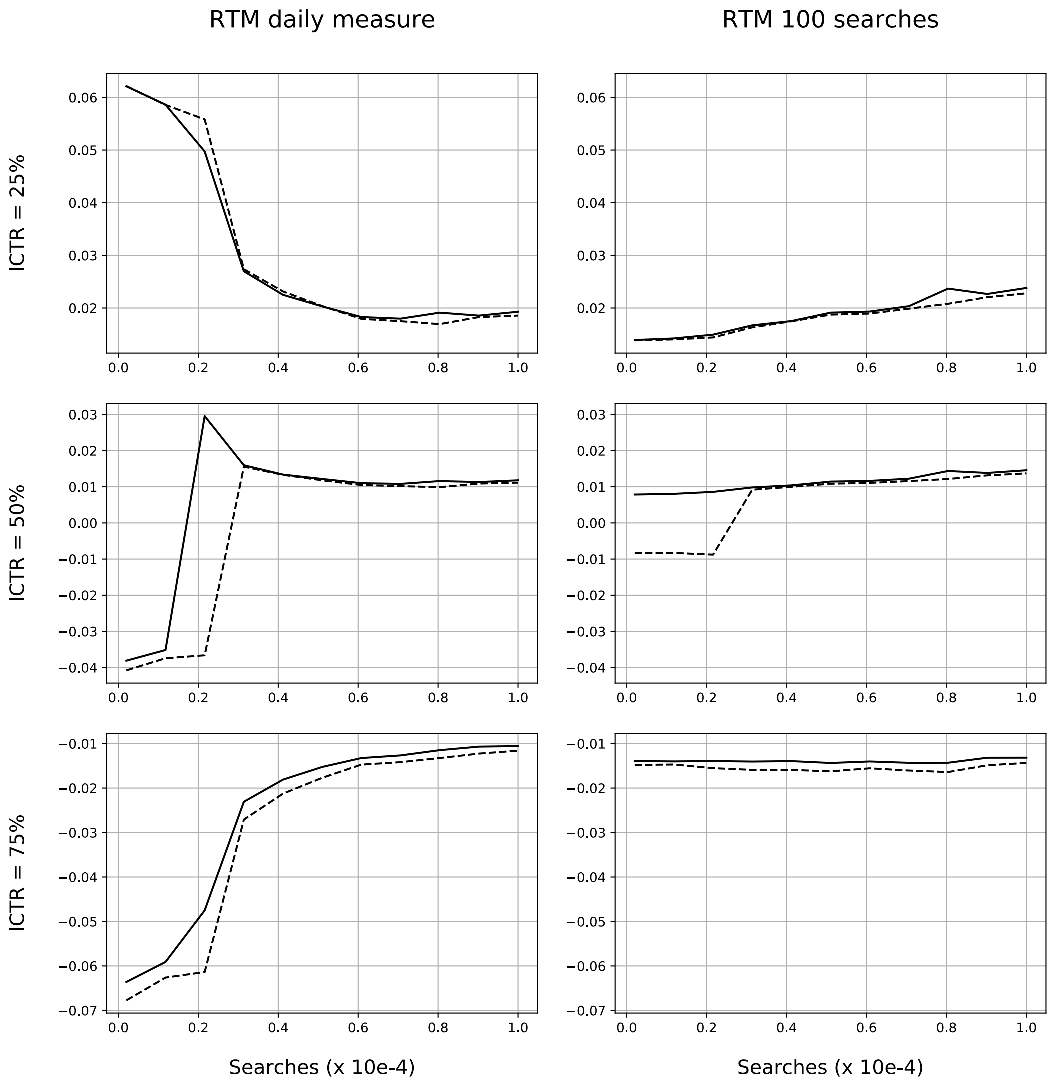



**Figure 2.27:** Left column: Estimated Regression to the mean effect when analysis is based on difference in the click through rate between first and last day. Right column: Estimated Regression to the mean effect for results in 2.5 of subsection2.4.2

## 2.7.5 Model Discussion

### 2.7.5.1 Proof of Proposition 1

**Confoundedness in types**

We start by proving the first case considered in Proposition 1. The first case deals with confoundedness in types under simultaneous unconfoundedness in age, which is modelled by $f(\mu_i, \mathrm{T}_i|s,h) = f(\mu_i|s,h) \times f(\mathrm{T}_i)$. The strategy of the proof is to rule out a convergent pattern in the type heterogeneity as data generating process for our data. We remind the reader that changes in expected values are assumed to be induced by first order stochastic dominance.

*Proof.* Consider two distinct values for the average user history $h$, $h^l$ and $h^s$, with $h^l > h^s$. Absent network effects, the difference between conditional initial qualities given $s_i = s$ and $\mathrm{T}_i = \mathrm{T}$ is given by $E[\mu|s, \Delta h]/(s \times \mathrm{T})^{\delta(h)}$, because $\mathrm{T}_i \perp h, s$. Assume that $\mathrm{E}[\mu|\bar{s}, \Delta h] < 0$ for some $s = \bar{s}$, i.e. queries with a longer average user history are on average of a lower type for $s = \bar{s}$. Since this implies that $E[\mu|\bar{s}, \Delta h]/(s \times \mathrm{T})^{\delta(h)} < 0 \quad \forall \mathrm{T}$, the integral of the latter expression over T, which yields $\mathrm{E}[ctr_i^1|s, \Delta h]$ must be negative. Hence, absent network effects, it is not possible that the model generates a pattern consistent with our data if $\mathrm{E}[\mu|s, \Delta h] < 0$ for some $s$. Now assume that $\mathrm{E}[\mu|s, \Delta h] >= 0 \quad \forall s$, and consider a convergent pattern, i.e. $\frac{\partial \mathrm{E}}{\partial s}[\mu|s, \Delta h] <= 0 \quad \forall s$. Any $\bar{s}$ and $\underline{s}$, such that $\bar{s} > \underline{s}$, implies that $\mathrm{E}[\mu|\underline{s}, \Delta h]/(\underline{s} \times \mathrm{T})^{\delta(h)} > \mathrm{E}[\mu|\bar{s}, \Delta h]/(\bar{s} \times \mathrm{T})^{\delta(h)} \quad \forall \mathrm{T}$. Consequently, we must have that $\mathrm{E}[ctr_i^1|\Delta h, \underline{s}] > \mathrm{E}[ctr_i^1|\Delta h, \bar{s}]$ for any $\bar{s}$ and $\underline{s}$ such that $\bar{s} > \underline{s}$. □

The proof rules out that a convergent pattern in the types could generate a divergent pattern in the initial quality. To see that a divergent pattern in types can in principle generate the data, note that a divergent pattern in types allows the possibility that $\mathrm{E}[\mu|\underline{s}, \Delta h]/(\underline{s} \times \mathrm{T})^{\delta(h)} < \mathrm{E}[\mu|\bar{s}, \Delta h]/(\bar{s} \times \mathrm{T})^{\delta(h)}$ for $\mathrm{T} < \mathrm{T}^\star$. I.e. if keywords are not too old, for example if all T in the age distribution are below $\mathrm{T}^\star$, the divergent pattern in types can be conserved. Note that for larger T the inequality reverses, which suggest that the divergent pattern is not stable. Furthermore, the expression $\mathrm{E}[\mu|s, \Delta h]/(s \times \mathrm{T})^{\delta(c)}$ illustrates that differences in average type vanish for larger T, hence, if the age distribution puts more weight on older queries, $\mathrm{E}[ctr_i^1|\Delta h, s]$ becomes smaller.

**Confoundedness in age**

The proof for age confoundedness is slightly more complex. In the case of type confoundedness, the proof is facilitated by the property that differences in quality caused by differences in the type of keywords vanish at the same speed, independently of the type of keywords. The absolute quality difference between two keywords $i$ and $j$ with $s_i = s_j = s$ and $\mathrm{T}_i = \mathrm{T}_j = \mathrm{T}$ caused by differences in types is given by $|\mu_j - \mu_i|/(s \times \mathrm{T})^{\delta(h)}$. By contrast, the absolute quality difference between two keywords $i$ and $j$ with $s_i = s_j = s$ and $\mu_i = \mu_j = \mu$ caused by differences in age is given by $\big((1-\mu) \times |\mathrm{T}_j^{\delta(h)} - \mathrm{T}_i^{\delta(h)}|\big)/\big(s \times \mathrm{T}_i \times \mathrm{T}_j\big)^{\delta(h)}$. The last expression reveals that the same age difference leads to a smaller quality difference if the corresponding product of the ages is larger.

*Proof.* Given $\bar{s}$ and $\mu$, the difference in the expected initial quality between $h^l$ and $h^s$ can be written as $\int \big(1 - \frac{1-\mu}{\bar{s} \times \mathrm{T}}\big) f(\mathrm{T}|\bar{s}, h^l) d\mathrm{T} - \int \big(1 - \frac{1-\mu}{\bar{s} \times \mathrm{T}}\big) f(\mathrm{T}|\bar{s}, h^s) d\mathrm{T}$. The latter expression is integrated over $f(\mu)$ to obtain $\mathrm{E}[ctr_i^1|\Delta h, \bar{s}]$. If $F(\mathrm{T}|\bar{s}, h^s)$ FODs $F(\mathrm{T}|\bar{s}, h^l)$, which implies $E[\mathrm{T}|\bar{s}, \Delta h] \leq 0$, because $1 - \frac{1-\mu}{\bar{s} \times \mathrm{T}}$ is a strictly increasing function in T, it follows that $\int \big(1 - \frac{1-\mu}{\bar{s} \times \mathrm{T}}\big) f(t|\bar{s}, h^l) d\mathrm{T} - \int \big(1 - \frac{1-\mu}{\bar{s} \times \mathrm{T}}\big) f(t|\bar{s}, h^s) d\mathrm{T} \leq 0 \quad \forall \mu$ and therefore $\mathrm{E}[ctr_i^1|\Delta h, \bar{s}] \leq 0$. Hence, absent network Effects, the model cannot generate a pattern consistent with our data if $E[\mathrm{T}|\bar{s}, \Delta h] \leq 0$ for some $s = \bar{s}$. Now assume that $\mathrm{E}[t_{i0}|s, \Delta h] >= 0 \quad \forall s$, and consider a convergent pattern, i.e. $\frac{\partial \mathrm{E}}{\partial s}[t_{i0}|s, \Delta h] \leq 0 \quad \forall s$. From $\int \left(1 - \frac{1-\mu}{s \times \mathrm{T}}\right) f(\mathrm{T}|s, h^l) d\mathrm{T} - \int \left(1 - \frac{1-\mu}{s \times \mathrm{T}}\right) f(\mathrm{T}|s, h^s) d\mathrm{T}$, it is easy to see that for larger $s$, the difference between the integrals reduces if $f(\mathrm{T}|s, h^l)$ and $f(\mathrm{T}|s, h^s)$ do not depend on $s$. Furthermore, reduction in the FOD of $F(\mathrm{T}|s, h^l)$ over $F(\mathrm{T}|s, h^s)$ also reduces the difference between the integrals for each $s$. Finally, if we replace $f(\mathrm{T}|s, h^l)$ and $f(\mathrm{T}|s, h^s)$ by $f^{'}(\mathrm{T}|s, h^l)$ and $f^{'}(\mathrm{T}|s, h^s)$, such that each $F^{'}$ FOD $F$ and $F^{'}(\mathrm{T}|s, h^l) - F^{'}(\mathrm{T}|s, h^s) \leq F(\mathrm{T}|s, h^l) - F(\mathrm{T}|s, h^s) \quad \forall s$, the difference between the integrals also reduces. Thus, if an increase in $s$ is accompanied by a decrease in the FOD of $F(\mathrm{T}|s, h^l)$ over $F(\mathrm{T}|s, h^s)$, which implies $\frac{\partial \mathrm{E}}{\partial s}[t_{i0}|s, \Delta h] \leq 0 \quad \forall s$, a divergent pattern cannot be generated, unless $F(\mathrm{T}|s, h^l)$ and/or $F(\mathrm{T}|s, h^s)$ increase with $s$, which we rule out. $\square$

#### 2.7.5.2 Discussion of the Model under Network Effects and no Heterogeneity

It is easy to show that the difference in quality between any two keywords $i$ and $j$ with $\overline{H_i} > \overline{H_j}$ and $S_{it} = S_{ij} = S$ and $\mu_i = \mu_j = \mu$ is positive $\forall S, \mu$. Furthermore it is a concave function in $S$ with an unique maximum $S^\star$, which is increasing for $S < S^\star$ and decreasing for $S > S^\star$. Keywords with a larger monthly popularity $s$ reach the region $s \times \mathrm{T} = S > S^\star$ earlier as a function of T. Thus, when integrating over T, keywords with a larger monthly popularity, $s$, will be assigned more weight over the domain $S > S^\star$. This is true $\forall \mu$. Therefore $\mathrm{E}[ctr_i^1|s, \Delta h]$ first decreases for larger $s$.

# Chapter 3

# Airbnb, Hotels, and Localized Competition[1]

**Bibliographic information: Schaefer, M., Tran, K.D., 2020. Airbnb, Hotels, and Localized Competition, DIW Discussion Paper No. 1889.**

## 3.1 Introduction

In fall 2007, the co-founders of the peer-to-peer short-term accommodation platform Airbnb hosted their first guest in their San Francisco apartment.[2] As of 2016, the platform listed three times as many rooms as the largest global hotel chain, Marriot International (Haywood *et al.*, 2017). This rapid rise was accompanied by controversial policy debates. On the one hand, consumers most likely benefit from Airbnb due to the increased variety in choices and the competitive pressure the online platform puts on traditional hotels, resulting in lower prices. On the other hand, the hotel industry argues that many of the hosts on Airbnb effectively run small hotels without facing the same level of regulation.[3] In this discussion, it is important to understand what is the actual nature of competition between Airbnb listings and hotels and to quantify how consumers profit from the presence of the peer-to-peer accommodation platform.

In this paper, we combine data on hotel and Airbnb demand in Paris in 2017 to estimate a discrete choice model of accommodation demand. Based on the estimates, we assess the nature of competition between Airbnb and hotels. In simulations, we next quantify how Airbnb affects consumer welfare as well as hotel revenues and demand.

One main contribution of our paper is that we can account for variation in demand within the

[1]This chapter is joint work with Kevin Tran. We thank the Fritz Thyssen Stiftung for their financial support. We also thank Reinhold Kessler and Matthias Hunold for kindly providing us with the hotel price data without which this project would not have been feasible. We are grateful to our supervisors Tomaso Duso and Hannes Ullrich for their helpful comments and guidance.

[2]See `https://news.airbnb.com/fast-facts/` (accessed: July 17, 2020).

[3]A 2017 study commissioned by the American Hotel & Lodging Association (AHLA) highlights the increasing importance of multi-unit hosts on Airbnb. This prompted AHLA President Katherine Lugar to call on Airbnb to "crack down on the illegal hotels that it facilitates." See `https://www.ahla.com/press-release/new-study-shatters-airbnb-homesharing-myth` (accessed: July 3, 2020).

city. Our data contains daily-level information on hotel and Airbnb demand in Paris. Furthermore, because we have information on the location of hotels and Airbnb listings within the city, we can allow for segmentation of demand not only along the type of accommodation but also its location. Consequently, we can assess how localized competition is within the city and include this localization in our counterfactual simulations accordingly. Our data suggest that local, short-term demand variation within the city plays an important role in the accommodation market. Prior research only uses variation across cities and time to identify their estimates (Zervas *et al.*, 2017; Farronato and Fradkin, 2018).

In 2017, Paris was the second most popular Airbnb destination in Europe, not far behind London.[4] Therefore, our analysis for Paris contributes important pieces of evidence for the European debate on the regulation of short-term accommodations. Existing research mostly focuses on the US.

Our results suggest that consumers perceive Airbnb listings and hotels as substitutes, but substitution is stronger within each of the two types of accommodation. In particular, consumers value hotels more highly than Airbnb listings, on average. Our findings also show that competition is localized. Consumers' preferences are correlated for accommodations in the same district. Consequently, accommodations in the same district are closer substitutes than accommodations in different districts. For example, we find that a ten percent price increase of five-star hotels leads to a one percent increase in demand for Airbnb listings in the same district, while the demand for accommodations in other districts increases by only 0.1 percent.

Based on the estimated demand parameters, we quantify the gain in consumer surplus from Airbnb in counterfactual simulations. Our results suggest an average gain in consumer surplus of roughly 4.3 million euro per night (or 31 euro per traveler). This gain in consumer surplus is a result of both the increased choice from the availability of Airbnb as well as of hotels charging lower prices with increased competition from Airbnb. Ignoring this price effect, we find that the average consumer surplus gain from Airbnb amounts to approximately 3.1 million euro per night. These gains in consumer surplus are heterogeneous over time: When hotels are near their capacity limits, consumer surplus gains from Airbnb total up to eight million euro per night (including the gain from lower hotel prices). Our results suggest that Airbnb is most beneficial to consumers during periods of high demand when hotels would charge higher prices absent Airbnb.

For our analysis, we combine data from three main sources. For Airbnb demand and prices, we use web scraped data obtained through AirDNA.[5] For hotel demand, we use a monthly survey conducted by the French National Institute of Statistics and Economic Studies (INSEE)[6] that asks a sample of French hotels for their daily occupancy rates. Because these data do not include hotel prices, we combine them with web scraped hotel prices from Booking.com, a hotel booking platform.[7] By combining these different data sets, we can estimate a joint model of hotel and Airbnb demand.

For the estimation, we define products based on their type (hotel or Airbnb listing), their quality, and the district in which they are located. In our main analysis, we propose a two-level nested logit model in which the first nesting tier represents the choice between different districts in the

[4]See `https://www.statista.com/statistics/957312/airbnb-leading-european-destinations/` (accessed: July 17, 2020).

[5]See `https://www.airdna.co/` (accessed: July 3, 2020).

[6]See `https://insee.fr/en/accueil` (accessed: July 3, 2020).

[7]The data collection was initially done as part of the work on Hunold *et al.* (2020). We are grateful to the authors for providing us with access to the data.

city and the second nesting tier represents the choice between hotels and Airbnb listings within a district. This nesting structure allows for correlated preferences by geography (districts) and accommodation type while remaining tractable.

Using the parameter estimates, we next simulate consumer welfare, hotel prices, and hotel demand in counterfactual scenarios. The main counterfactual we are interested in is a world in which Airbnb is absent. To simulate this counterfactual, arguably the biggest hurdle is simulating hotel prices. The main problem is that hotels' profit maximization most likely needs to account for the capacities of nearby hotels as well as their own capacity constraints. Analytically solving a price equilibrium in this setting would be complex. Therefore, we propose a heuristic approach, to solve for the equilibrium numerically. The basic idea is to iteratively assign predicted demand to hotels up to their capacity constraints and repredict the remaining excess demand. At the same time, hotel prices that are consistent with the assigned demand need to be found.

Our work is most closely related to the work by Zervas *et al.* (2017) and Farronato and Fradkin (2018). Using data from Texas, Zervas *et al.* (2017) find that an overall increase in Airbnb supply reduces hotel revenues by approximately ten percentage points in their sample, with lower-priced and non-business hotels being disproportionately strongly affected. They show that the revenue impact is mainly driven by a reduction in prices as opposed to a reduction in the number of rooms booked.[8]

Closest to our work, Farronato and Fradkin (2018) employ a random coefficient discrete choice model and estimate the demand and supply parameters using data for several US cities. Consistent with our results, they find that consumers profit most from Airbnb in locations and at times when hotels are capacity constrained. Because they use data on the city level, they identify capacity constraints when demand for the entire city increases (e.g. for New Year's Eve in New York City). In our granular data, we show that capacity constraints in individual districts in the city are actually binding a lot more frequently than those for the entire city. Thus, combined with the localized level of competition that our results suggest, accounting for variation within a city seems important to analyze the accommodation market.

Furthermore, our research is related to a growing body of literature studying competition between peer-to-peer platforms and incumbent firms. Seamans and Zhu (2014) and Kroft and Pope (2014) analyze the impact of Craigslist on incumbent industries. Aguiar and Waldfogel (2018) assert that streaming affects music sales using Spotify data. Cohen *et al.* (2016) and Lam *et al.* (2020) study consumer surplus from Uber, while Hall *et al.* (2018) address the question of Uber's complementary to public transport systems. Cramer and Krueger (2016) assess the relative efficiency of Uber drivers compared to taxi drivers.

More broadly, our paper is also related to the literature studying externalities related to the platform economy. One research strand studies how Airbnb affects housing markets (Horn and Merante, 2017; Koster *et al.*, 2018; Garcia-López *et al.*, 2020; Barron *et al.*, 2020; Duso *et al.*, 2020; Valentin, 2020). Basuroy *et al.* (2020) analyze Airbnb's impact on local neighborhoods. Our paper is also broadly related to the emerging literature on online platforms and surge pricing (Cachon *et al.*, 2017; Castillo, 2019; Guda and Subramanian, 2019) as we show that Airbnb is particularly valuable for consumers when hotels charge higher prices.

The remainder of the article is organized as follows. In Section 3.2, we describe some institutional details of the Parisian accommodation market and introduce the various data sources that we use

[8]In a similar exercise, Neeser (2015) uses data from three Nordic European countries. Unlike Zervas *et al.* (2017), the author does not find a statistically significant impact of Airbnb on hotel performance.

in more detail. In Section 3.3, we present our empirical strategy by showing selective descriptives, introducing our model, providing some estimation details, and finally discussing identification of the key model parameters. In Section 3.4, we report the estimated parameters as well as demand elasticities. In Section 3.5, we outline our simulation procedure and present the results of the counterfactual analysis. Section 3.6 concludes.

## 3.2 Institutional Details and Data

In our analysis, we assess the competition between Airbnb and hotels using data for the city of Paris in 2017. Therefore, in this section, we first briefly present some facts about the accommodation industry in Paris. Subsequently, we introduce our main data sets in some detail. Finally, we describe how we define products for our analysis.

### 3.2.1 The Short-Term Accommodation Market in Paris

According to Mastercard's Global Destination Cities Index, Paris is the city with the third largest number of international visitors worldwide in 2017.[9] Only Bangkok and London attracted more international travelers. In the 2019 edition of the same report, Paris even surpassed London and was the second most visited city in the world.[10] Accordingly, Paris is also the second most popular Airbnb destination in Europe, not far behind London, in 2017.[11] Therefore, studying the competition between Airbnb and hotels in Paris amounts to studying it in one of the most important short-term accommodation markets not only in Europe, but the world.

Paris has introduced regulations that limit the number of days an apartment can be rented out during a calendar year. Since October 2017, the City of Paris requires hosts of entire apartments on Airbnb to obtain and display a registration number. Since January 2020, hosts of entire homes can no longer rent out their apartments for more than 120 days in a year.

### 3.2.2 Data

Our goal is to estimate a joint model of Airbnb and hotel demand in Paris. Therefore, the main necessary components for our analysis are demand and prices for hotels and Airbnb listings. To obtain this information, we combine three different data sets that we describe in the following.

**Hotel Demand Data** To assess hotel demand, we use data provided by INSEE. Each month, INSEE surveys a random sample of French hotels and asks for information regarding occupancy and origin of guests.[12] Our main question of interest concerns the number of booked rooms for

[9] See `https://mastercardcontentexchange.com/newsroom/press-releases/2018/big-cities-big-business-bangkok-london-and-paris-lead-the-way-in-mastercards-2018-global-destination-cities-index/` (accessed: July 17, 2020).

[10] See `https://newsroom.mastercard.com/wp-content/uploads/2019/09/GDCI-Global-Report-FINAL-1.pdf` (accessed: July 17, 2020).

[11] See `https://www.statista.com/statistics/957312/airbnb-leading-european-destinations/` (accessed: July 17, 2020).

[12] The population of hotels is first stratified by region, star rating, and type (chain or independent). Within each stratum, random samples are drawn. INSEE reports that a sample of approximately 12,000 out of 18,000 hotels is contacted in the national sample. See `https://insee.fr/en/metadonnees/source/operation/s1480/processus-statistique` (accessed: July

each day of the survey month. Based on this self-reported daily occupancy, we create our measure of hotel demand. For each hotel, the data also include the star rating, the number of available rooms, and the district.[13] Based on the hotel responses, INSEE imputes the occupancy for non-respondents as well as non-sampled hotels. Consequently, the data contain (imputed) quantity information for the entire population of hotels.

**Hotel Price Data** The hotel quantity data from INSEE lacks an important component for demand estimation: hotel prices. Thus, we use web scraped hotel price data from the hotel booking platform Booking.com. For each night in 2017, the data include the prices for each hotel available on Booking.com two weeks, one week, and one night prior to the night of interest. The prices reported are those shown as a result of a search for a one-night stay of two persons. If there are multiple room categories, the reported price is that for the cheapest available category.[14] To match this price information to the hotel quantity data, we use hotel names. This match is possible for 983 of the 1,597 hotels in the quantity data.[15] Due to the occasional instability of the web scraper, the number of observed hotels varies. Therefore, we exclude dates from the analysis for which we do not observe prices for at least 65 percent of the total number of rooms. As a result, 232 dates remain in our analysis.

**Airbnb Data** For our analysis, we require data not only on the supply of Airbnb listings in Paris, but also on the corresponding demand. Obtaining a measure of Airbnb supply is possible by web scraping the Airbnb website. More specifically, by searching the website for listings in Paris without restricting the search to any specific dates, all listings that are available for booking at some date in the future can be found. However, obtaining a measure for Airbnb demand is more difficult. The reason is that Airbnb does not display whether a listing was booked for a given date. In principle, for each listing a calendar is shown that indicates whether it is available for booking on a given date or not. However, if a listing shows as unavailable, this can mean that it is booked for that date or that the host blocked it for private reasons.

Therefore, we use web scraped Airbnb data available from AirDNA.[16] The data include all listings that were available for booking in Paris in 2017. The advantage of the AirDNA data is that they include information on whether a listing was available, blocked, or booked on each date. This information is imputed by AirDNA. They are able to do so because they gathered data from Airbnb when the website itself still displayed whether an unavailable listing was booked or blocked. They use this information to predict whether an unavailable listing in the present data is most likely booked or blocked. While imperfect, this measure for Airbnb demand is the best available that we are aware of with the exception of data obtained from Airbnb directly. In addition to the imputed booking status, the data include the price, the location, and further listing-specific characteristics that are observable on the Airbnb website.

For the analysis, we also need a measure of listing quality to account for the impact of quality on demand. While hotels have a natural quality measure (their star rating), this is not the case for Airbnb listings. Therefore, to define quality categories for Airbnb accommodations, we follow

16, 2020).

[13]The exact address of the hotels is also available but since the data are classified as sensitive, we are prohibited from geocoding the addresses.

[14]See Hunold *et al.* (2020) for more details on the data gathering procedure.

[15]We compare matched and non-matched hotels in Appendix 3.7.1.

[16]See `https://www.airdna.co/` (accessed: July 3, 2020).

Farronato and Fradkin (2018). We regress Airbnb prices on time and listing fixed effects. Subsequently, we divide the estimated listing fixed effects by the quartiles of their distribution and categorize each Airbnb listing into one of the four groups according to their estimated fixed effect. As a result, we create four Airbnb quality categories.

### 3.2.3 Product Definition

In principle, we observe individual transactions for Airbnb listings. For hotel demand, we have the occupancy of each hotel on each date. Therefore, it is conceivable to conduct our demand estimation using a disaggregated discrete choice model. A major advantage of such a disaggregated analysis would be that it would allow us to take into account the specific location of each product. However, while we do have the precise location for each Airbnb listing, we do not observe this for hotels due to the sensitivity of the data. Instead, we only know in which district each hotel is located.[17] Therefore, if we were to estimate demand on a disaggregated level, we would nevertheless only be able to use geographic information on the district level.

Another potential advantage of disaggregated estimation is that variation in consumers' choice sets can help to empirically identify parameter estimates. However, we do not observe the time at which each consumer booked their accommodation. Therefore, we do not observe if consumer choice sets have changed due to accommodations having been booked. Thus, including variation in choice sets in the model would require additional, unverifiable assumptions.

Finally, one disadvantage of disaggregated estimation is its immense computational burden. In particular, with potentially large choice sets that often occur in the context of online platforms, disaggregated discrete choice estimation can quickly become computationally infeasible.

Due to these concerns, we decided to aggregate the data for our analysis. For the aggregation, we define a product as a combination of type (Airbnb listing or hotel), quality category, and district. As a result, we circumvent the problems of disaggregated analysis discussed above while still being able to consider geographic heterogeneity within the city (on the district level).

As quality categories, we use the star ratings for hotels and the fixed effects categorization due to Farronato and Fradkin (2018) and described in Section 3.2.2 for Airbnb listings. As prices, we use the average transaction price for each product for a given date of stay. Note that these prices are missing for a fraction of hotels. As a result, the average hotel prices are based on only those hotels for which price data is available. However, when calculating the aggregate demand for each product, we use the entire sample of hotels. If we used only those hotels for which we have prices, the demand for hotels with no price data would wrongly be assigned to the outside good. Therefore, an implicit assumption in our aggregation is that the average prices of hotels with and without price information are comparable.

With this aggregation procedure, our analysis is based on a maximum of 180 possible products per market.[18] The product market shares are defined as the ratio of the number of booked rooms

[17] In principle, it would be possible to geocode each hotel in the hotel price data based on its name. This would yield information on the location of each hotel for which we have prices. While this geocoding based on names would probably not work perfectly throughout, we expect that it should work well for most hotels. However, since INSEE does not permit geocoding of the hotel addresses, it is unethical to do so in a workaround and would likely violate the INSEE use agreement. Furthermore, this approach would mean that we can only use 983 out of 1,597 hotels (those for which we can successfully match the price to the quantity data).

[18] These 180 products result as a combination of 20 districts and nine type-quality combinations (four Airbnb quality categories and five possible hotel stars).

over the market size for a particular night. We treat each night as a separate market. This assumption is necessary, because, as discussed earlier, the hotel quantity data do not allow us to distinguish between individual bookings.

# 3.3 Empirical Strategy

In this section, we discuss our empirical strategy. We begin by presenting selected descriptives that highlight particularities of the Parisian short-term accommodation market that are important for identification and estimation. Next, we describe the model on which the estimation is based. Subsequently, we set forth some estimation details. Finally, we discuss some issues concerning the identification of the model parameters.

## 3.3.1 Descriptives

For a better understanding of the data and to highlight issues of importance for estimation and identification, we first present some descriptives. In particular, we show how Airbnb and hotel supply changes over time. Further, we document evidence for the importance of localized competition in the short-term accommodation market in Paris. The results presented in this subsection motivate our choice of the demand model and the identification strategy that we discuss in subsequent subsections.[19]

The main components to any demand estimation are supply, demand, and prices. Therefore, we begin our descriptive analysis by examining the supply side of the short-term accommodation market in Paris. Figure 3.1 presents the number of rooms available in Paris for each date in 2017, split by hotels and Airbnb listings. Airbnb listings make up approximately 40 percent of the overall room capacity in Paris, on average. However, while hotel room supply is almost constant over time, the supply of Airbnb listings fluctuates both in the short- and long-runs. Around October 2017, the number of Airbnb listings in the city decreases substantially. This decrease is likely due to the introduction of a mandatory registration number display for Airbnb listings in Paris starting in October 2017.[20]

Turning to demand over time, Figure 3.2 presents the number of rooms occupied for both types of accommodation. Contrasting the supply patterns, demand for hotels varies more strongly. Usually, hotel demand is higher during weekdays than weekends. Airbnb demand, on the other hand, is more stable in the short run. Despite this difference in variance, demand for Airbnb and hotels appear to move roughly in sync. This joint movement suggests that demand for the two types of accommodation is, at least, not completely independent from one another.

Next, we consider descriptives that illustrate the relationship between Airbnb and hotel prices. Figure 3.3 shows the average daily prices for the two types of accommodation.[21] The average hotel price exhibits larger variation and follows the demand patterns shown in Figure 3.2. In comparison, the variation observed for Airbnb prices is negligible. This figure suggests that there are fundamental differences in price-setting behavior of hotels and Airbnb listings. This insight

[19] We present summary statistics by hotel and Airbnb quality category in Appendix 3.7.1.

[20] This decrease in Airbnb listings is arguably exogenous to demand for accommodations in Paris. However, focusing exclusively on the period around the policy change is not feasible for our analysis, because most of the observed transactions take place before October 2017.

[21] Note that, for hotels, we only include the days for which we have sufficient price information.

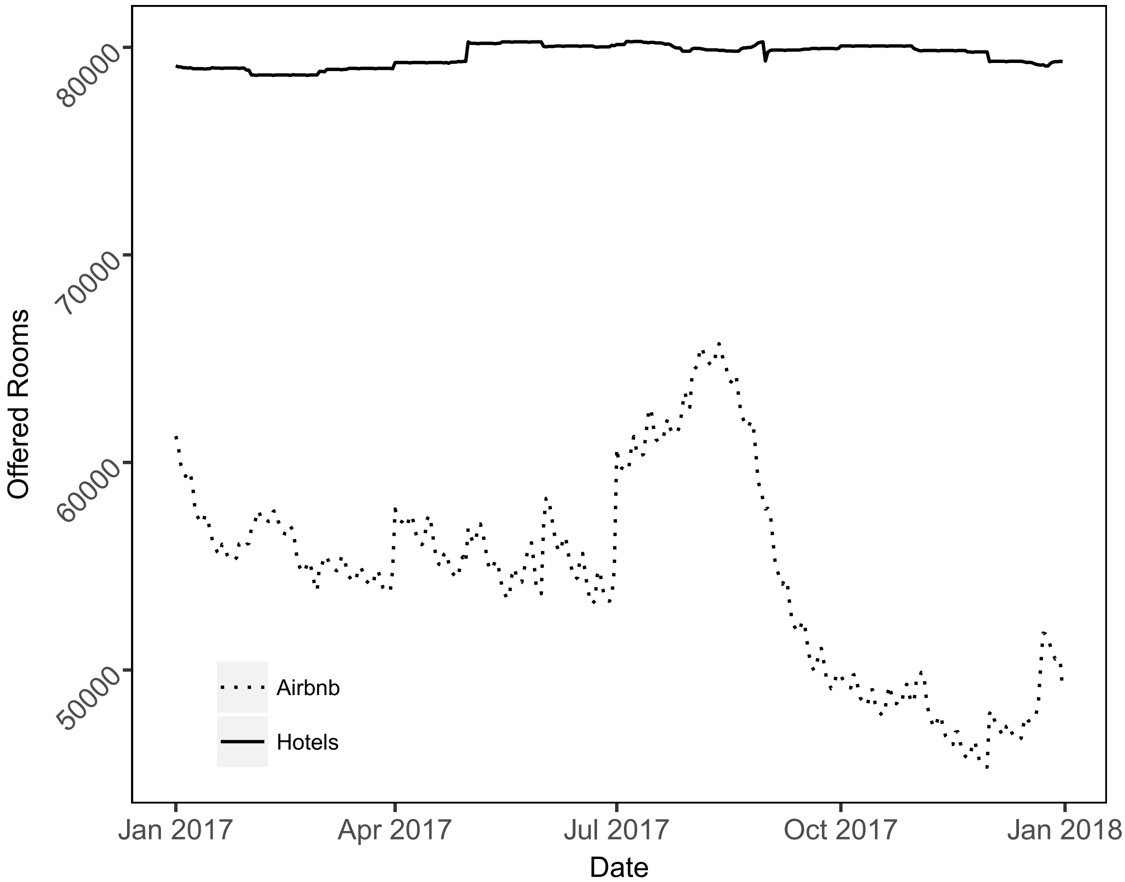



**Figure 3.1:** Offered rooms by date, split by hotels and Airbnb listings

is important, especially when we simulate hotel and Airbnb prices in counterfactual scenarios in later sections.

The next set of descriptives combines an analysis of demand and price setting. Figure 3.4a illustrates how hotel prices correlate with the average occupancy ratio (i.e. the number of booked rooms over the total number of available rooms). For exposition, this figure shows the relationship for four-star hotels in the first district. However, this pattern can be observed more generally for other types of accommodation and other districts as well. In general, hotel prices increase with higher occupancy ratios. However, the relationship appears to be non-linear. Specifically, when hotels reach their capacity limits, prices increase more sharply. When simulating hotel price setting, we need to take this pattern into account.

A similar graphical analysis for Airbnb prices and occupancy rates is not informative. The reason is that Airbnb capacity never reaches its limits. However, Figure 3.3 suggests that Airbnb prices do not react as strongly to changing demand as hotel prices do. Instead, Figure 3.1 indicates that Airbnb listings may rather react to changing demand by adjusting supply. To assert whether this relationship is true, we propose to analyze the joint movement of the deviation of Airbnb occupancy and supply from long-term trends. To calculate long-term Airbnb occupancy and supply, we calculate a 14-day moving average for both. Next, for each date, we calculate the deviation of actual occupancy and supply from these moving averages. These deviations can be interpreted as short-term fluctuations in Airbnb occupancy and supply, net of long-term trends. Figure 3.4b

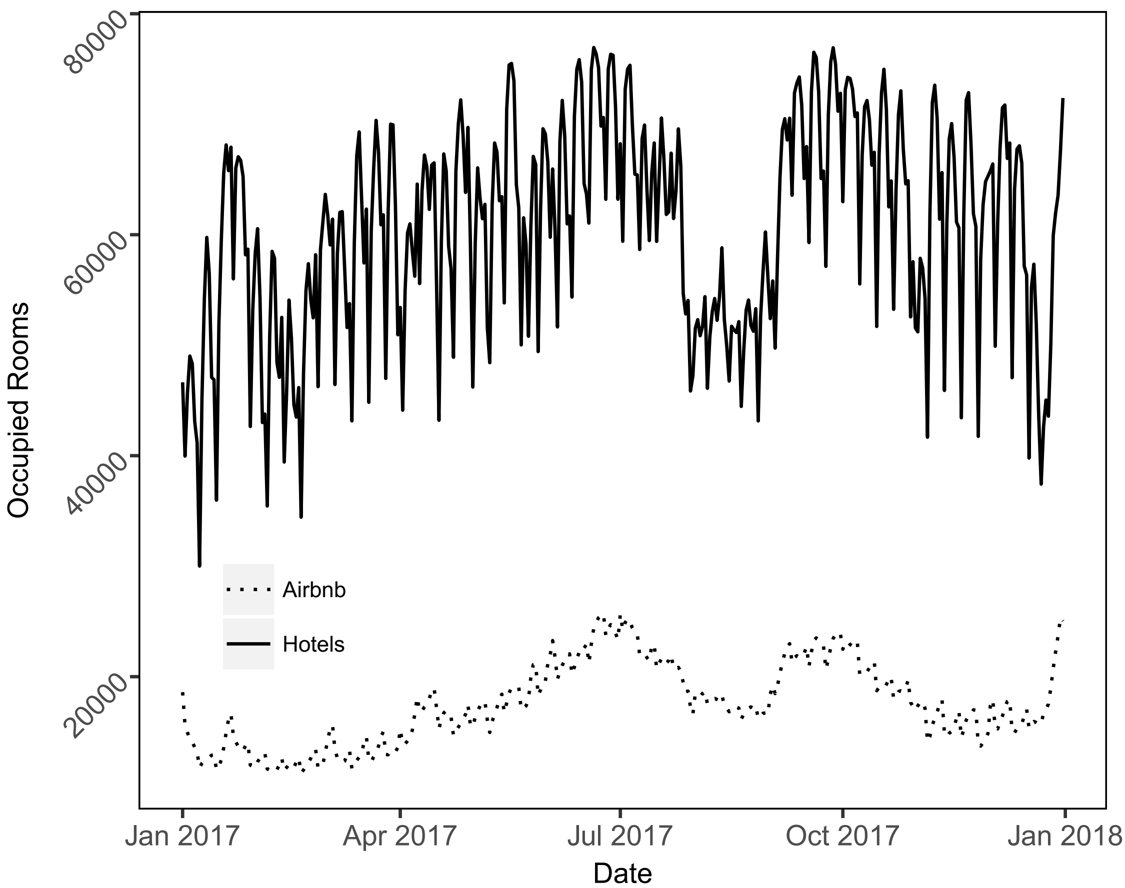



**Figure 3.2:** Occupied rooms by date, split by hotels and Airbnb listings

shows a scatterplot of this residual occupancy and supply with a linear fit. Each point represents one date. The positive slope of the linear fit suggests that as occupancy (i.e. demand) for Airbnb increases in the short run, the supply of Airbnb also tends to be larger. Thus, this figure is in line with Airbnb hosts flexibly reacting to changing market situations and, consequently, Airbnb supply being more flexible than hotel supply.

Figure 3.4a suggests that capacity constraints of hotels play an important role in the market for short-term accommodations. In the final set of descriptives, we assess the prevalence of situations in which these capacities are constrained. Figure 3.5 shows the city-wide hotel occupancy ratio over time in the solid black line. The horizontal dashed line marks an occupancy ratio of 90 percent which we define as a high occupancy state. There are several dates during the year when the black line is above this 90 percent cut-off. Farronato and Fradkin (2018) find that consumers profit most from the presence of Airbnb during these high occupancy states. However, in Figure 3.5, we further document that localized high occupancy states occur far more frequently than city-wide high occupancy states. The dotted line shows the share of districts in which hotel occupancy exceeds 90 percent at each date. For many dates when the city-wide hotel capacity does not seem to be constrained, it actually is in individual districts of the city. This insight implies that demand variation is not distributed uniformly across the city. If, additionally, consumer preferences are correlated within districts in the city, ignoring this within-city variation in occupancy rates, prices, and demand, likely distorts the estimation. One main advantage of our data compared to prior

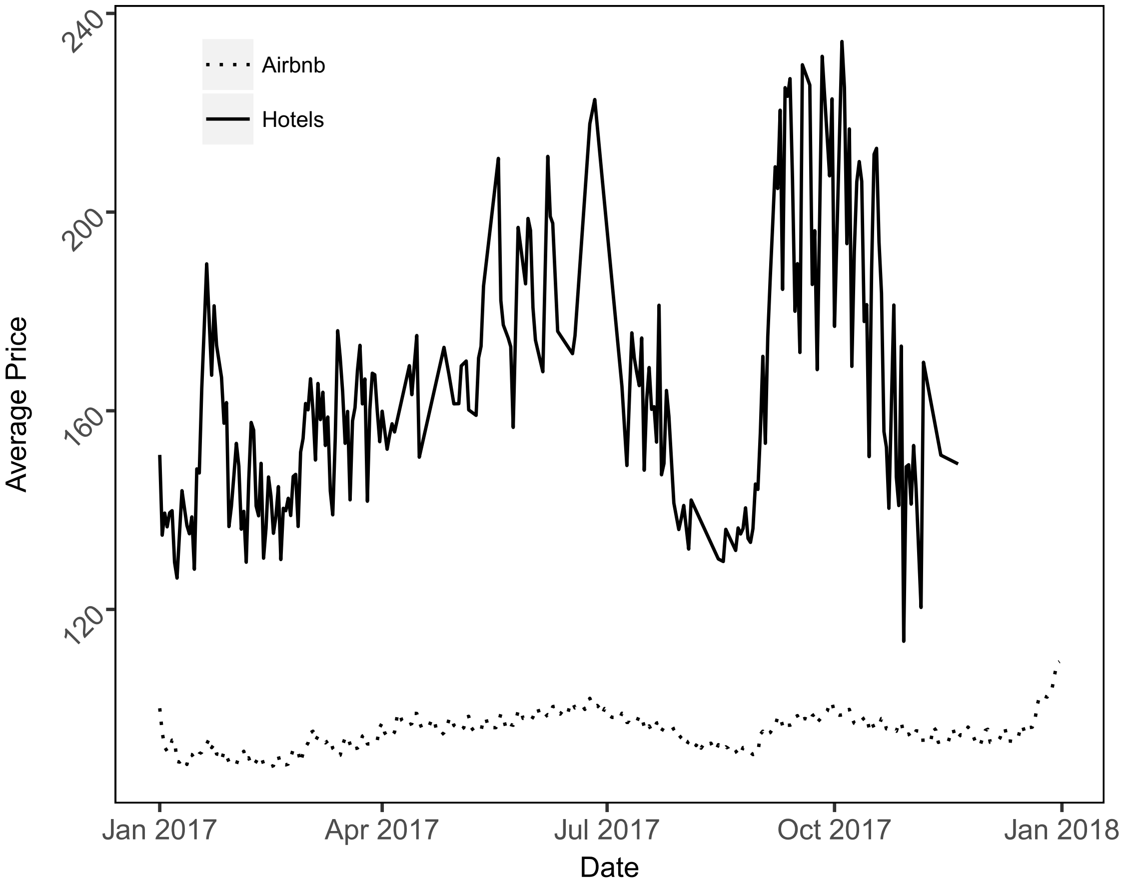


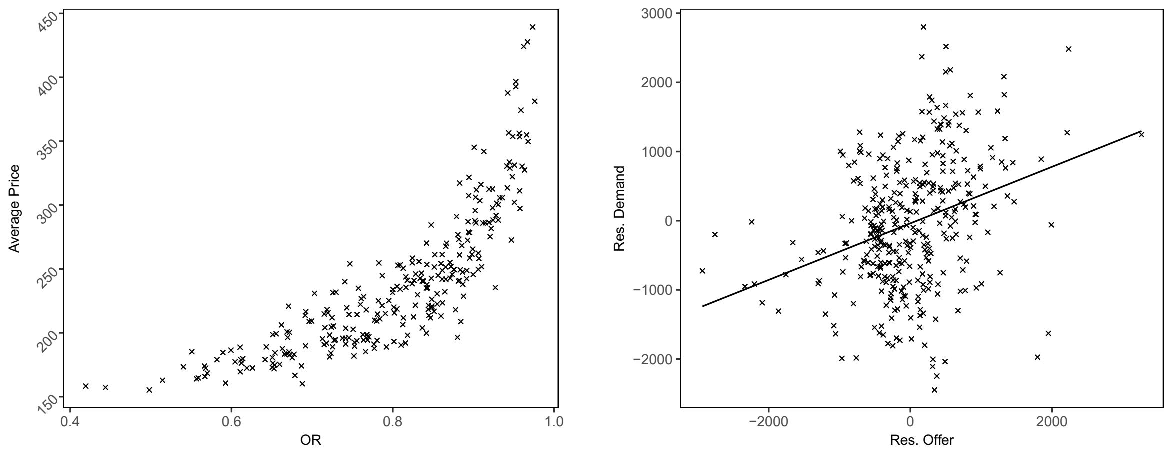

**Figure 3.3:** Average hotel and Airbnb prices over time

research is that it allows us to exploit the local variation observed in the data.

The capacities of Airbnb never really reach their limits. However, specifically when hotels are capacity constrained, Airbnb listings can serve as a valuable alternative for consumers (Farronato and Fradkin, 2018). Figure 3.6 shows the relationship between the occupancy ratio of hotels and Airbnb listings in an exemplary district. Note that the two axes have different scales. The vertical dashed line marks a 90 percent occupancy ratio of hotels. Each point represents one date. This figure provides two main insights. First, there is a positive correlation between hotel and Airbnb occupancy. Figure 3.2 already suggests such a correlation for absolute demand for hotels and Airbnb listings; Figure 3.6 confirms this pattern for occupancy ratios as well. Second, when hotel occupancy is high, Airbnb occupancy seems to be disproportionately higher as well. The cloud of points in the top right corner of the plot suggests that substitution from hotels to Airbnb listings might be more relevant when hotels are capacity constrained. This insight indicates that these high occupancy states are particularly important when considering how consumers profit from the presence of Airbnb.

### 3.3.2 Model

The evidence in Section 3.3.1 suggests that demand for short-term accommodations in Paris varies not only over time but also between different districts in the city. Furthermore, while some substitution between Airbnb and hotels takes place (compare Figure 3.6), demand does seem to be




**(a)** Prices and occupancy ratios (OR) for four-star hotels in the first district

**(b)** Short-term supply and occupancy for Airbnb listings net of long-term trends with linear fit

**Figure 3.4:** Occupancy rate and hotel and Airbnb behavior

somewhat segregated between the two types of accommodation. Therefore, we propose to base the demand estimation on a two-level nested logit model. The first nesting level represents a choice between different districts of the city. Figure 3.5 suggests that much of the variation in demand over time is not distributed uniformly across the city but is localized in individual districts. Defining the choice between districts as the first nesting level allows for preferences to be correlated among alternatives within each district and, therefore, can capture the important role that within-city geography plays in the demand for accommodations. The second nesting level represents the choice between Airbnb listings and hotels within each district. While the descriptive evidence suggests that geography is a major factor in the demand for accommodations, this second nesting level allows for some additional segmentation of demand along the types of accommodation.

We next describe the model more formally. We define each date as an individual market $t$. Let $J_t$ denote the set of available choices in market $t$. This choice set is mostly constant over time and consists of the products described in Section 3.2.3 (combinations of district, accommodation type, and quality category). However, in some markets, some Airbnb types are missing in certain districts. Each consumer chooses either one of the products in $J_t$ or the outside good. Let $d = 0, ..., D$ denote the district nests, where $d = 0$ is the outside good. Further, define $H_d$ as the subnests within each nest $d$.

The two-level nested logit model implies the following choice probability for product $j$, located in subnest $h_d$ in nest $d$:

$$P_{jt} = P_{jt|h_d} \times P_{h_d|d} \times P_d \tag{3.1}$$

$P_{jt|h_d}$ denotes the choice probability of product $j$ conditional on the consumer choosing subnest $h_d \in H_d$. $P_{h_d|d}$ denotes the probability of choosing subnest $h_d$ conditional on choosing nest $d$. $P_d$ denotes the probability of choosing nest $d$. The (conditional) probabilities in Equation (3.1) can be rewritten as (we follow the notation introduced by Verboven (1996)):

$$P_{jt} = \frac{e^{\delta_{jt}/(1-\sigma_1)}}{e^{I_{hdt}/(1-\sigma_1)}} \times \frac{e^{I_{hdt}/(1-\sigma_2)}}{e^{I_{dt}/(1-\sigma_2)}} \times \frac{e^{I_{dt}}}{\sum_{d=0}^{D} e^{I_{dt}}}\,, \tag{3.2}$$

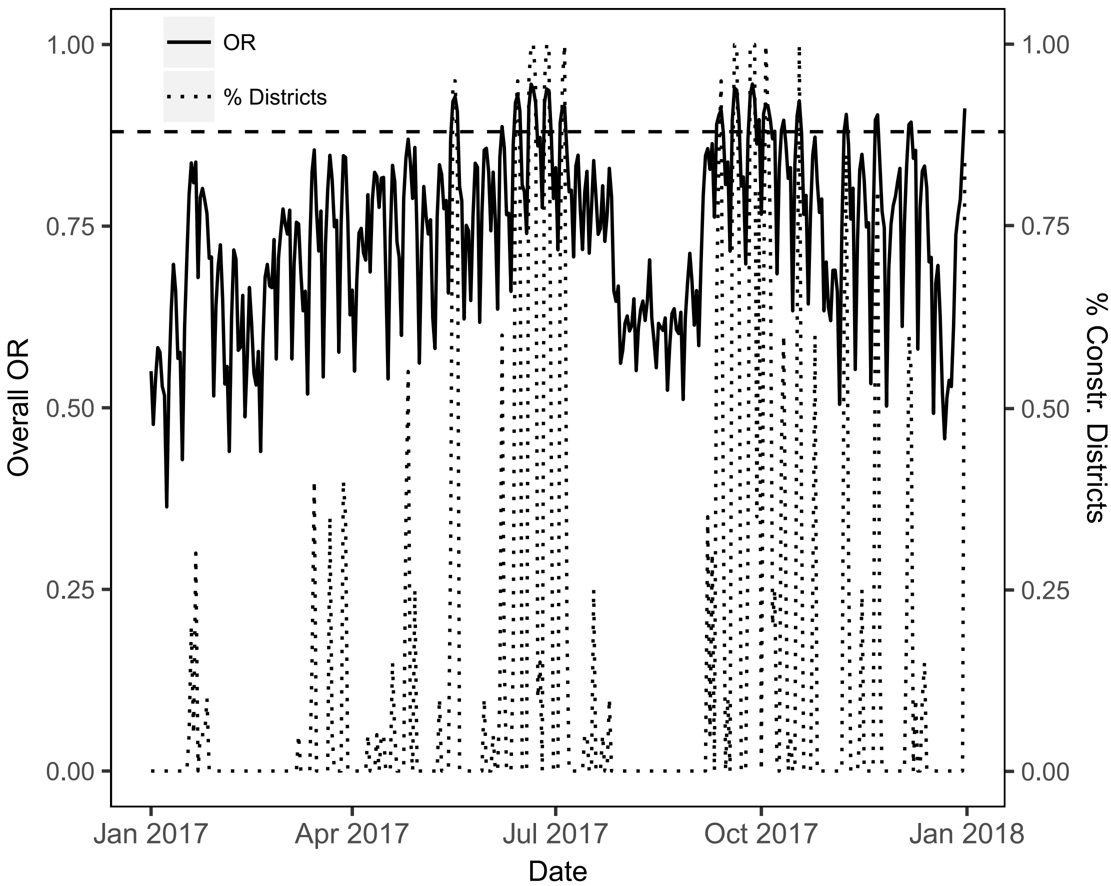



**Figure 3.5:** City-wide hotel occupancy ratio (OR) and percentage of districts in which hotel occupancy ratio is above 90 percent. The horizontal dashed line marks an occupancy ratio of 90 percent.

where $I_{hdt}$ and $I_{dt}$ represent the expected utility from choosing an alternative in subnest $h_d$ and nest $d$, respectively. These expressions are given by:

$$\begin{aligned} I_{h_d t} &= (1-\sigma_1) ln\Big(\sum_{j \in h_d t} e^{\delta_{jt}/(1-\sigma_1)}\Big) \text{ and} \\ I_{dt} &= (1-\sigma_2) ln\Big(\sum_{h_d \in H_{dt}} e^{I_{h_d t}/(1-\sigma_2)}\Big). \end{aligned} \tag{3.3}$$

The parameters $\sigma_1$ and $\sigma_2$ govern the correlation of preferences between alternatives in the same subnest and nest, respectively. The parameters must fulfill $0 \leq \sigma_2 \leq \sigma_1 \leq 1$ to be consistent with random utility maximization of the consumers (McFadden *et al.*, 1977).

A larger value of $\sigma$ implies stronger correlation in the unobserved preference components. The inequalities therefore imply that the preferences across alternatives in the same subnest must be weakly more strongly correlated than preferences across alternatives in the same nest, but different subnests. Note that, in principle, the parameters $\sigma_1$ and $\sigma_2$ can vary across different nests. A fully flexible model would require estimating 60 parameters: 20 parameters for each upper level nest (one for each district) and 40 parameters for each lower level subnest (one for hotels and one for Airbnb in each district). To facilitate estimation, we restrict $\sigma_1$ and $\sigma_2$ to be identical across nests

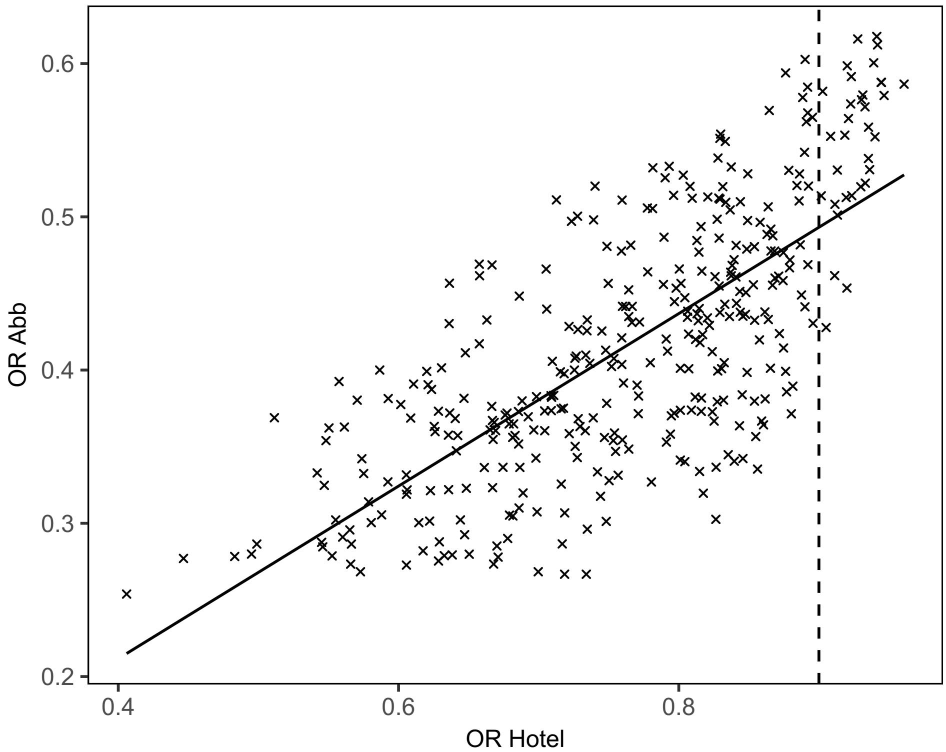



**Figure 3.6:** Occupancy ratio (OR) of hotels and Airbnb listings in first district

and subnests.

$\delta_{jt}$ is the mean utility of alternative $j$ in market $t$, which we define as follows:

$$\delta_{jt} = x_{jt}\beta - \alpha p_{jt} + \epsilon_t + \epsilon_{jt} \, . \tag{3.4}$$

$x_{jt}$ is a vector of observable product characteristics, $p_{jt}$ denotes the price, $\epsilon_t$ is a market-specific fixed effect and $\epsilon_{jt}$ is a product-specific time varying error term that is potentially correlated with the price. We assume that vector $x_{jt}$ is uncorrelated with $\epsilon_{jt}$. In our analysis, $x_{it}$ contains quality category fixed effects, a constant, and, in some specifications, time fixed effects, high-capacity districts fixed effects, and the log of exogenous Airbnb supply. All of these variables are fixed in the short term and cannot react to changing demand. Therefore, we believe that the assumption that $x_{jt}$ is exogenous, is reasonable.

### 3.3.3 Estimation

We estimate the parameters $\alpha$, $\beta$, $\sigma_1$, and $\sigma_2$ using the market share inversion technique outlined in Berry (1994) and Berry *et al.* (1995). Verboven (1996) derives the inverted market shares for a two-level nested logit model, which yields the following estimation equation:

$$ln(s_{jt}/s_{0t}) = x_{jt}\beta - \alpha p_{jt} + (\sigma_2 - \sigma_1) ln(s_{jt}/s_{hdt}) + \sigma_2 ln(s_{jt}/s_{dt}) + \epsilon_t + \epsilon_{jt} \tag{3.5}$$

$s_{jt}$ is the share of product $j$ in market $t$. $s_{0t}$ denotes the share of the outside good. $s_{hdt}$ is the cumulative share of products in subnest $h_d$ and $s_{dt}$ is the cumulative share of products in nest $d$.

To determine the market shares, we require an estimate of the market size, i.e. a measure of the number of people looking for short-term accommodation in Paris. We compute the market size using worldwide Google trends data for the keyword “hotels Paris” and “Airbnb Paris.” The data span the last four weeks of 2016 and all weeks in 2017. We add up both trends to obtain a measure for the total search interest for both accommodation types. We normalize the aggregate trend by setting its average value equal to the average combined number of hotel and Airbnb rooms in 2017. As a result, one unit of the combined Google trends data corresponds to 1100 searches.

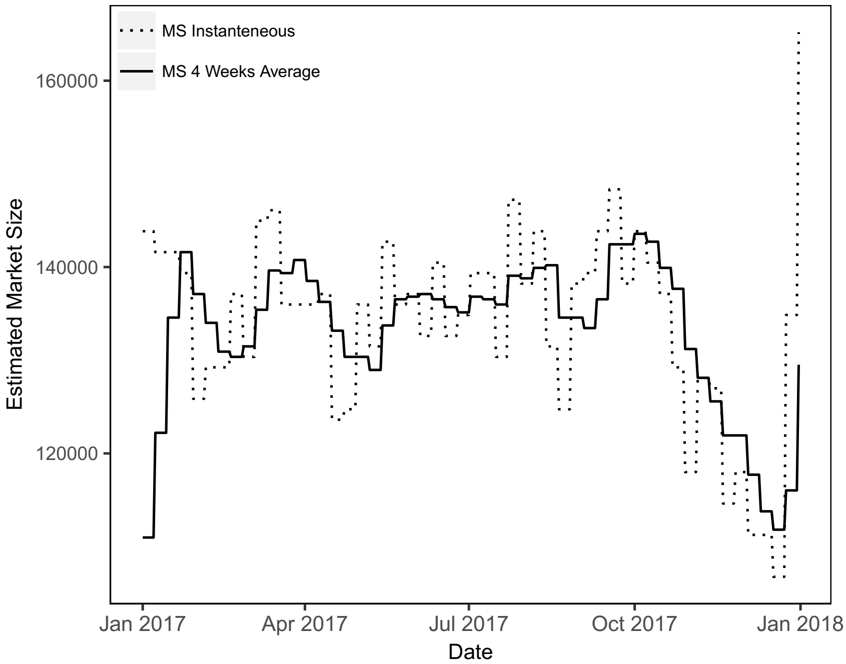



**Figure 3.7:** Market size (MS) measures

The dashed line in Figure 3.7 shows our measure for the search activity. Our measure suggests that, on average, there are 134,000 daily searches globally for either Airbnb listings or hotels in Paris in 2017. The measure ranges from 107,000 to 165,000. Based on these search numbers, we calculate our main market size measure for each date as the average of the searches in the four weeks prior to that day. The solid line in Figure 3.7 presents this measure. This moving average captures the idea that travelers plan their trip in advance. However, our results are robust to using the number of searches at each date as the market size (i.e. the dashed line in Figure 3.7).[22]

### 3.3.4 Identification

There are two main concerns for the identification of the model parameters. The first is the standard concern in demand estimation: There are three endogenous variables on the right hand side of Equation (3.5): the price and the two share terms. Therefore, we need at least three instrumental variables to consistently estimate Equation (3.5). The second concern is that our model does not capture differences in the capacities of the products. In its basic form, Equation (3.5) does not account for the number of individual Airbnb listings or hotels that are underlying each of the aggregated products. Ignoring changes in these numbers might bias our results.

[22]Table 3.8 in Appendix 3.7.4 reports the results of this robustness check.

**Instrumental variables** One main concern regarding the endogeneity of prices and market shares are time-varying localized demand shocks. As discussed in Section 3.3.1, the different districts are subject to short-term demand variation that is heterogeneous in location and intensity over time.

Demand shocks enter $\epsilon_{jt}$ and are positively correlated with prices. Without appropriate instrumental variables, this will result in an underestimation of the magnitude of the true price parameter. Note that the localized demand shocks will also lead to biased parameter estimates of $\sigma_1$ and $\sigma_2$ since any unobserved demand shock that affects $s_{jt}$ will necessarily also affect the share terms on the right hand side of Equation (3.5). Therefore, we require instrumental variables that are independent from local, short-term demand variation and correlated with prices and market shares.

Our instrumental variable strategy relies on the insight that the price setting behavior of a given accommodation is dependent on the room capacity of its nearby rivals. *Ceteris paribus*, higher room capacity of nearby competitors should reduce the ability to charge higher prices.

Therefore, we propose to construct instruments based on the room capacity of similar products (i.e. the same accommodation type and quality category) in other districts. In this measure, further away competitors should be weighted less. To formalize this idea, consider a product $j$, in market $t$. Let $A_{jt}$ be the set of products of the same type and quality in market $t$. With 20 districts, this set can contain a maximum of 20 elements. We construct the instrument for product $j$ as:

$$z_{jt} = \sum_{\substack{l \in A_{jt} \\ l \neq j}} \frac{c_{lt}}{e_{lj}^2}\,, \tag{3.6}$$

where $c_{lt}$ is the capacity of product $l$ and $e_{lj}$ is the Euclidean distance in kilometers between products $j$ and $l$.[23] Thus, the instrument is the weighted sum of the rival capacity, with weights that are inversely related to the distance to the rival.

For $z_{jt}$ to be a valid instrument, it must be uncorrelated with short-term variation in demand. Figure 3.1 suggests that this is likely the case when calculating the measure for hotels, as hotel demand is almost constant over time. However, for Airbnb supply, the descriptives presented in Section 3.3.1 suggest that the capacity does respond to demand shocks. Therefore, we need a measure of the *exogenous* supply of Airbnb listings to use in the calculation of Equation (3.6) for Airbnb listings.

For this purpose, we propose to predict the Airbnb supply for each product using explanatory variables that are plausibly exogenous with regard to short-term variation in demand. More specifically, we predict Airbnb supply based on a quartic time trend and a measure of leisure-related outgoing car traffic from Paris. The quartic time trend flexibly captures changes in Airbnb supply in the city due to long-term trends that should be independent from short-term and localized variation in demand. To obtain a measure of leisure-related outgoing car traffic, we use data that measure outgoing car traffic at main highway exits around Paris.[24] To distinguish between business traffic and leisure-induced traffic, we construct a measure of weekend and holiday traffic. On the day before the weekend or the holiday (i.e. on Fridays for each weekend), we take the amount of outgoing traffic and subtract the average amount of outgoing traffic during the weekdays of the

[23]We use the centroids of each district to calculate the distances.

[24]The data are available at `https://opendata.paris.fr/explore/dataset/comptages-routiers-permanents/information/?disjunctive.libelle&disjunctive.etat_trafic&disjunctive.libelle_nd_amont&disjunctive.libelle_nd_aval` (accessed: July 20, 2020).

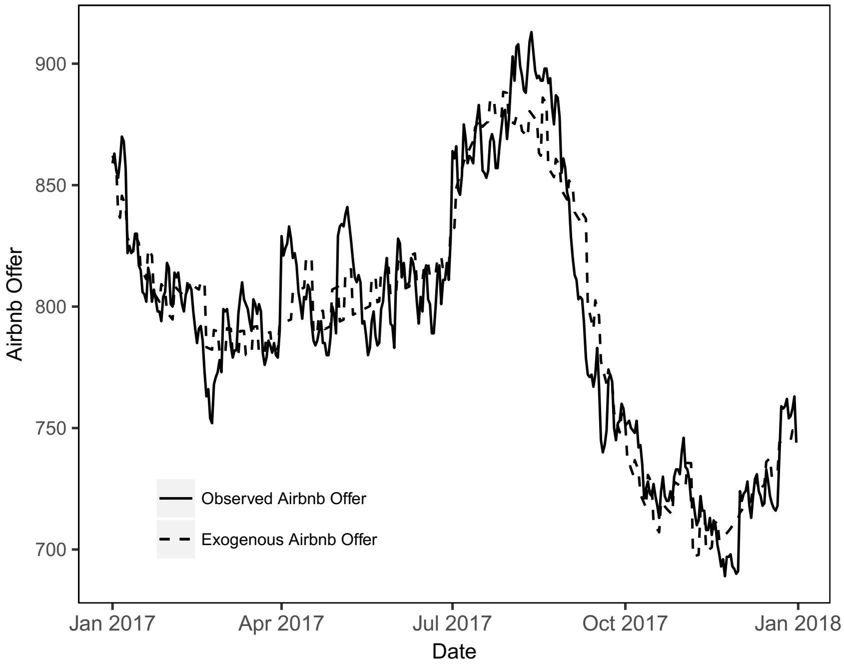



**Figure 3.8:** Observed and exogenous Airbnb offer for selected Airbnb product

preceding week.[25] We interpret this excess traffic as travelers leaving the city for the weekend or for the holidays. For the estimation, we assign this excess value to the entire respective weekend or holiday. We let the measure be zero for weekdays. Therefore, weekday variation in Airbnb supply is only explained by long-term trends. The procedure is applied for each Airbnb product separately to allow for product-specific variation of the exogeneous offer. Figure 3.8 shows the observed and predicted Airbnb offer for a selected Airbnb product. The deviations from the long-term trend are caused by the variation in the traffic data. Having predicted Airbnb supply using the quartic time trend and the leisure-induced outgoing car traffic, we use these predicted Airbnb numbers to calculate Equation (3.6) for all Airbnb products.

With this procedure, we obtain measures of Equation (3.6) that only capture same-type capacities in the city. However, the capacities of products of the other type with similar quality are likely important for price setting and market shares as well. To capture this, we assign the values calculated in Equation (3.6) to products of similar quality but different type in the same district. For example, say we calculate $z_{jt}$ for the category-four Airbnb product in district one. We would then assign this value also to four- and five star hotels in district one.[26] As a result, each observation has two instruments: The capacity of same type and quality products in other districts as well as the capacity of different type and similar quality products in other districts. Finally, we interact each of these two instruments with a fixed effect for hotels. Through this interaction, the instruments can affect hotels differently than Airbnb listings. This choice is motivated by the analysis in Section 3.3.1, which reveals differences in the response of Airbnb and hotels to

[25] The summer holiday peak in Airbnb offer is an exception. We do not use the outgoing traffic to capture this peak. Instead, we interact the quartic time trend with a dummy variable that captures the period of the summer holidays.

[26] Because we have four Airbnb and five hotel quality categories, we assign the values of the highest Airbnb category to both the four as well as the five star hotels.

short-term variation in demand.

We generate an additional instrument using the capacity of accommodations of the other type in the same district. For example, for Airbnb listings in a given district, we use the total hotel capacity in that district. For hotels, we use the total Airbnb capacity. Again, we interact this instrumental variable with a fixed effect for hotels.

As noted above, the capacity of hotels is nearly constant over time. To generate temporal variation in the instruments based on hotel capacity, we interact all the above instruments with the market size. The identifying assumption is that the market size is independent of the exact location of demand shocks. The intuition behind this assumption is that the total market size is an aggregate of the location-specific demand. Therefore, variation in demand for specific locations in the city does not affect total demand to travel to Paris much and is offset on average. The descriptive analysis in Figure 3.5 shows that periods of district-level hotel capacity constraints becoming binding are much more frequent than periods of city-wide hotel capacities being at their limits. This insight is in line with our assumption.

Another issue regarding the validity of the market size as an instrument is its relationship with the overall price level in the city. The market size reflects the number of individuals who are considering travel to the city at a specific point in time. Long-term variation in average accommodation prices across different tourist destinations most likely affect the market size for each of these destinations. However, for us to use the market size as an instrument, we only require that it is uncorrelated with short-term variation in prices. If most individuals traveling to Paris are somewhat inflexible in the dates of travel (this would be the case e.g. for business travelers or tourists planning their trips in advance), then short-term price variation should not affect the total market size.

The reverse mechanism that hotels anticipate periods of high demand and set higher prices accordingly is more problematic. Consumers' responses to relative price differences within the city should remain unaffected by the overall shift in prices. This suggest that the variation in relative shares within the city as a response to variation in price differentials within the city should allow us to consistently estimate the price parameter.[27]

**Accounting for capacity** Our model does not capture changes in the number of hotels or Airbnb listings underlying each of the aggregated products. Accounting for differences in these numbers might be relevant for our results. However, the descriptives presented in Section 3.3.1 suggest that the Airbnb and hotel supplies behave fundamentally differently. Consequently, accounting for differences in the actual number of rooms underlying the products requires different approaches for each accommodation type. For hotels, the main concern is that differences in hotel capacities might be driving part of the differences in market shares. Hotels with low capacity tend to have lower market shares. Therefore, market shares might reflect capacity constraints rather than consumer preferences. Because hotel capacities do not change much in our sample, we propose

[27]We perform a simulation study where we assess the validity of the instrument. In our simulation, four products (which can be understood as hotels) adjust their prices in response to overall market size and product-specific shocks. Consumers' choice probabilities follow a simple logit model. Other than prices, products are differentiated by one exogenous observed characteristic. As an instrument we use the average value of the exogenous product characteristic of rivals (classical BLP instrument) and interact it with the market size. The resulting 2SLS estimator produces a consistent estimator. Details of the simulation study are presented in Appendix 3.7.2, where we also compare the performance of the proposed instrumental variable with the classical BLP instrument without interaction with the market size.

to capture these differences by including fixed effects for hotels in districts with a high capacity of four- and five-star hotels.[28] We define a district as a high-capacity district if it contains more than 2,000 four- and five-star hotel rooms.[29] This criterion applies to six districts. We include a separate fixed effect for each of these districts. We let the fixed effects be zero for all Airbnb listings.[30]

The issue is different for Airbnb capacities. Because the number of Airbnb listings changes over time, a fixed effect would not be sufficient to capture the effect of these changes. Instead, we follow Ackerberg and Rysman (2005) and include the log of the Airbnb capacity (we add one to allow for capacities of zero) in the estimation. Section 3.3.1 shows that Airbnb capacity likely reacts to short-term changes in demand. Therefore, we only include the log of the *exogenous* Airbnb capacity that we also use to calculate Equation (3.6) for Airbnb listings. For hotels, we let this variable be zero. These additional variables allow us to account for changes in the market share that might be driven by unobserved changes in the underlying number of rooms and listings. A caveat of these additional variables is that they lack a clear structural interpretation. Therefore, we also report results without these additional variables.[31]

# 3.4 Results

## 3.4.1 Parameter Estimates

Table 3.1 shows the result of estimating different specifications of Equation (3.5). Estimation is performed using two-stage least squares, the reported standard errors are heteroskedasticity robust.[32]

Column (1) reports the results for a simple logit estimation without instruments. The positive price coefficient suggests that it suffers from an upward bias due to unobserved factors increasing both demand as well as prices. The estimated quality category fixed effects are also counter-intuitive: For instance, five-star hotels are valued lower than two-star hotels.

[28] We only focus on four- and five-star hotels because these make up a majority of overall hotel demand. High-quality hotels with large capacities are more likely to host events such as conferences. Furthermore, four- and five-star hotels also reflect differences in hotel capacity across districts well. In districts in which there is a large capacity of four- and five-star hotels, overall room capacity tends to be higher. As a potential alternative to our approach, de Palma *et al.* (2007) propose a method to account for capacity constraints in demand estimation. However, their method is only developed for a simple multinomial logit. Extending this method to more flexible demand models is an interesting avenue for future research.

[29] Effectively, a district is defined as high capacity if it has a four- and five-star hotel capacity of 2,629 or more hotels which is the smallest of the district capacities above 2,000. A capacity of 2,629 corresponds approximately to the 75 percentile of the distribution of district-wide four- and five-star hotel capacities.

[30] In Figure 3.14 of Appendix 3.7.1, we show the distribution of hotel and Airbnb capacity over districts.

[31] An alternative way to assess the sensitivity of our results to the varying Airbnb offer is to restrict the estimation sample to only contain observations between February 1st and July 1st. As Figure 3.1 reveals, the Airbnb offer varies less during this period. We present the results of the corresponding analysis in Table 3.9 of Appendix 3.7.4. Our main results are not affected when restricting the sample period.

[32] We report the first-stage F-statistics in Table 3.7 of Appendix 3.7.3. We use the same set of instruments for each endogenous variable. The F-statistics suggest that the instruments are not weak.

**Table 3.1:** Parameter estimates

| | (1) Logit | (2) Logit (IV) | (3) NL | (4) NL | (5) NL |
|---|---|---|---|---|---|
| Price | 0.002*** | −0.008*** | −0.006*** | −0.006*** | −0.005*** |
| | (0.0001) | (0.0004) | (0.0003) | (0.0003) | (0.0003) |
| $\sigma_1$ | | | 0.758*** | 0.657*** | 0.717*** |
| | | | (0.017) | (0.014) | (0.018) |
| $\sigma_2$ | | | 0.256*** | 0.228*** | 0.207*** |
| | | | (0.010) | (0.009) | (0.006) |
| Airbnb category 2 | 0.290*** | 0.501*** | 0.068*** | 0.132*** | −0.019* |
| | (0.041) | (0.044) | (0.021) | (0.023) | (0.011) |
| Airbnb category 3 | 0.421*** | 0.924*** | 0.280*** | 0.379*** | 0.078*** |
| | (0.040) | (0.047) | (0.027) | (0.027) | (0.018) |
| Airbnb category 4 | 0.261*** | 1.586*** | 0.774*** | 0.920*** | 0.456*** |
| | (0.042) | (0.067) | (0.044) | (0.044) | (0.038) |
| 1-star hotel | −0.539*** | −0.148*** | 0.698*** | 0.619*** | 2.553*** |
| | (0.048) | (0.052) | (0.034) | (0.030) | (0.066) |
| 2-star hotel | 0.740*** | 1.319*** | 0.968*** | 1.032*** | 2.712*** |
| | (0.041) | (0.048) | (0.025) | (0.025) | (0.093) |
| 3-star hotel | 2.067*** | 2.922*** | 1.463*** | 1.674*** | 3.232*** |
| | (0.041) | (0.055) | (0.041) | (0.040) | (0.121) |
| 4-star hotel | 1.644*** | 3.182*** | 1.822*** | 2.045*** | 3.505*** |
| | (0.043) | (0.075) | (0.050) | (0.050) | (0.126) |
| 5-star hotel | −0.210*** | 4.326*** | 3.206*** | 3.482*** | 4.468*** |
| | (0.066) | (0.176) | (0.114) | (0.111) | (0.165) |
| Log Airbnb offer | | | | | 0.351*** |
| | | | | | (0.016) |
| Constant | −6.470*** | −6.031*** | −4.157*** | −4.307*** | −6.319*** |
| | (0.039) | (0.044) | (0.045) | (0.070) | (0.128) |
| Market FEs | *N* | *N* | *N* | *Y* | *Y* |
| N | 33,504 | 33,504 | 33,504 | 33,504 | 33,504 |
| Adjusted $R^2$ | 0.377 | 0.133 | 0.652 | 0.700 | 0.862 |

Notes: The lowest Airbnb quality category serves as the base category. *, **, *** indicate statistical significance at the ten, five, and one percent level, respectively. F-statistics are shown in Appendix 3.7.3.

Columns (2) and (3) show the results for a simple logit specification and the two-level nested logit model of Equation (3.5) when using the instruments presented in Section 3.3.4. Column (4) includes market-specific fixed effects. Column (5) further includes the natural logarithm of the exogenous Airbnb offer and district-specific fixed effects for all hotels located in a district with a high capacity of four- and five-star hotels. In all of these instrumented specifications, the price coefficient is statistically significantly negative and appears robust across specifications.

Generally, consumers seem to value hotels more highly than Airbnb listings. Our results suggest that Airbnb listings of the highest quality are valued similarly to two-star hotels. Note that in column (5), the Airbnb and hotel fixed effects are not readily comparable because we include high-capacity district fixed effects that apply to hotels only while the log Airbnb offer applies only to Airbnb listings. The coefficient for the logarithm of the exogenous Airbnb offer is statistically significantly different from zero. Ackerberg and Rysman (2005) propose a structural interpretation for this coefficient: The result implies that consumers value higher availability of Airbnb listings because, as a result, accommodations are more likely to be in line with their geographic preferences.[33]

The estimate for $\sigma_1$ suggests strong taste correlation between products in the same subnest. Recall that each subnest is an accommodation type (hotel or Airbnb listing) in a given district. Furthermore, the estimate for $\sigma_2$ indicates taste correlation between products in the same district. However, the correlation between products in the same nest (i.e. in the same district) is smaller than for products in the same subnest (e.g. Airbnb listings in the same district). These results are in line with localized competition between Airbnb and hotels. However, localized competition within hotels and Airbnb listings in the same district is stronger than the competition between hotels and Airbnb listings in the same district.

### 3.4.2 Elasticities

Table 3.2 reports the average estimated own- and cross-price elasticities for different Airbnb and hotel quality categories based on the estimates of column (5) in Table 3.1.[34] Demand seems to be elastic for all quality categories except for the two lowest Airbnb quality categories. The own price elasticity tends to increase with the quality. Generally, the demand for hotels seems to be more elastic than the demand for Airbnb listings. However, the two highest Airbnb quality categories have comparable own-price elasticities to those of three- and four-star hotels, respectively.[35]

Given the estimates of $\sigma_1$ and $\sigma_2$, cross-price elasticies are highest for products in the same subnest. Cross-price elasticies have intermediate values for products that only share the same nests but are in different subnests. Finally, cross-price elasticities are lowest for products in different nests.

For example, Table 3.2 reveals that a one percent price increase for five-star hotels leads to an average 1.2 percent increase in demand for hotels in the same district. The same price increase results in a demand increase of only 0.1 percent for Airbnb listings in the same district. The demand for accommodations in other districts only increases by 0.01 percent following the price increase.

[33]For better exposition of the main results, we do not report the estimated coefficients for the high-hotel-capacity district fixed effects described in Section 3.3.4. All of these coefficients are statistically significantly positive, as expected. The high-hotel-capacity fixed effects lack the same structural interpretation of the logarithm of the Airbnb offer because high capacity does not necessarily imply large variety in the location of hotels within a district: In the extreme, a single hotel could make up the majority of the hotel capacity in a district.

[34]The formulas for the own- and cross-price elasticities are presented in Appendix 3.7.5.

[35]We report the semi-elasticities in Appendix 3.7.6. As opposed to the elasticities, the own-price semi-elasticities tend to decrease with the quality category, although the pattern is not as clear cut as for the own-price elasticities.

**Table 3.2:** Average estimated demand elasticities

| | | | Cross-price elasticities | | |
|---|---|---|---|---|---|
| Type | Category | Own-price elasticity | Same district & Type | Same district | Other district |
| Airbnb | 1 | -0.5095 | 0.2548 | 0.0091 | 0.0007 |
| | 2 | -0.9583 | 0.1744 | 0.0057 | 0.0004 |
| | 3 | -1.3400 | 0.2387 | 0.0077 | 0.0006 |
| | 4 | -2.4661 | 0.6033 | 0.0178 | 0.0014 |
| Hotels | 1 | -1.3769 | 0.0669 | 0.0053 | 0.0005 |
| | 2 | -1.6417 | 0.1310 | 0.0109 | 0.0011 |
| | 3 | -1.6144 | 0.6378 | 0.0538 | 0.0054 |
| | 4 | -2.6882 | 0.7415 | 0.0629 | 0.0060 |
| | 5 | -7.4321 | 1.2201 | 0.1111 | 0.0120 |

## 3.5 Airbnb, Consumer Welfare, and Hotel Performance

In this section, we assess the impact of Airbnb on hotel revenues and consumer welfare. There are two main channels through which Airbnb benefits consumers. First, Airbnb increases the choice variety. Second, Airbnb's presence in the market is likely to reduce hotel prices. To quantify these effects of Airbnb, we analyze counterfactual scenarios in which Airbnb does not exist and compare them to a scenario with Airbnb. In order to determine hotel prices in these counterfactual scenarios without Airbnb, we need to address the supply side of the market more closely. In the remainder of this section, we first present our supply-side model. Based on this model, we can simulate the price response of hotels to Airbnb leaving the market. This allows us to compare the loss in consumer surplus when Airbnb leaves the market *and* hotels adjust prices to a scenario in which hotels do not adjust prices when Airbnb leaves the market. Additionally, modelling the price response of hotels in the counterfactual scenario where Airbnb is not present in the market allows us to assess the impact of Airbnb on hotel revenues.

The standard approach to quantifying the impact of Airbnb would be to simulate counterfactual scenarios without Airbnb and compare the simulated outcomes to the observed outcomes in the data. However, note that the observed data are characterized by location-specific demand shocks that our model does not account for.[36] Hence, our simulation analysis is not able to incorporate these demand shocks. Therefore, we do not compare simulated outcomes (which do not incorporate unobserved short-term localized demand variation) with observed outcomes (which do include unobserved short-term localized demand variation). Instead, we simulate market outcomes not only for the scenarios without Airbnb but also for a benchmark scenario including Airbnb. Subsequently,

[36] In fact, our identification strategy builds on finding instruments that are orthogonal to location-specific demand shocks. Another approach would be to estimate market-district fixed effects directly. These fixed effects would capture short-term district-specific variation in demand. Subsequently, we could use the estimates in the counterfactual simulations. Unfortunately, this approach appears to be too demanding of the data. Particularly, note that with market-district fixed effects, the number of parameters grows with the sample size, which gives rise to dimensionality issues.

we compare hotel performance and consumer welfare in the scenarios without Airbnb to this benchmark scenario. Comparing simulation outcomes can be seen as studying the impact of Airbnb while abstracting from location-specific demand shocks.

To simulate scenarios without Airbnb, we only require a model to simulate hotel behavior absent Airbnb. However, to simulate the benchmark scenario, we also need to capture Airbnb supply. Therefore, we first discuss how we incorporate Airbnb supply in our simulations. Then, we describe how we model hotel supply. Subsequently, we present how well our simulation of the benchmark scenario fits the observed data. Finally, we report the results of our simulations.

### 3.5.1 Airbnb Supply

To include Airbnb supply in our simulation, we propose to use the exogenous measure of Airbnb supply discussed in Section 3.3.4. This exogenous Airbnb offer allows us to include Airbnb supply in a direct and simple way and is consistent with abstracting from local, short-term variation in demand in our simulations.[37] For Airbnb prices in our simulation of the benchmark scenario, we use the observed Airbnb prices. The descriptive analyses in Section 3.3.1 suggest that Airbnb supply reacts to short-term changes in demand by adjusting offered quantities rather than prices. Therefore, short-term variation in demand should not affect these observed Airbnb prices much.

### 3.5.2 Hotel Supply-Side Model

In this section, we describe our approach to model hotels' equilibrium prices. Conventionally, discrete choice models are extended to include the supply side by assuming a mode of competition between firms and deriving firms' profit-maximizing first order condition. This first order condition can then be used to derive marginal costs based on estimated demand elasticity and observed prices (Berry, 1994; Verboven, 1996). Together with the parameter estimates, these estimated marginal costs then allow for simulating prices in counterfactual scenarios.

In our application, this standard approach is problematic for two main reasons. First, as Figure 3.5 suggests, hotel capacity constraints play an important role in this market. Discrete choice models, however, typically assume unconstrained capacities on the supply side. When estimating the demand model, these constraints are somewhat alleviated because the demand parameters fit the theoretical market shares to the observed market shares, which, by definition, adhere to the capacity constraints. However, when simulating counterfactual situations, nothing restricts the resulting market shares from exceeding the actual capacities.

Second, optimal pricing of hotels likely requires internalization not just of the own capacities but also those of the competitors. This cross-dependency between competitors makes finding an analytical solution for the market equilibrium complex.

Therefore, we propose an approach to simulating hotel prices and quantities in counterfactual equilibria using a somewhat more heuristic, iterative approach. There are several components to our approach. The first component we need are hotels' optimal prices as a function of their occupancy rates. While in reality, this function might also rely on competitors' occupancy rates,

[37]Farronato and Fradkin (2018) model Airbnb hosts as price-takers who increase supply in response to higher prices. Since our model does not incorporate localized demand shocks, Airbnb supply predicted by an entry model would not adapt to demand shocks and therefore remain largely constant within a location across time. Therefore, the insight from modeling Airbnb entry in our context would be limited.

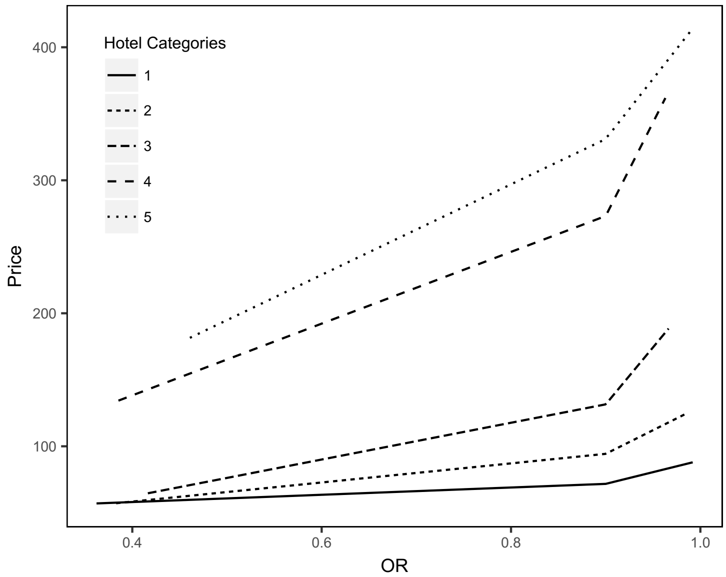



**Figure 3.9:** Estimated pricing function for different hotel types in an exemplary district

we propose to use the observed, empirical relationship between hotel prices and own occupancy as a reduced form approximation. The underlying assumption is that prices in the data are the result of optimal pricing by hotels. Assuming that the relationship between occupancy rates and optimal prices remains fixed without Airbnb, we can use the observed empirical relationship between prices and occupancy rates to simulate optimal hotel prices.

To estimate the function of optimal hotel prices dependent on occupancy ratios, we regress the observed prices on the corresponding occupancy ratios. Motivated by the insights from Figure 3.4a, we assume a kinked linear relationship between occupancy ratio and prices, with a kink at an occupancy ratio of 90 percent. We estimate a separate pricing function for each hotel product. Figure 3.9 shows the resulting estimated optimal prices as a function of the occupancy ratio for different hotel types in an exemplary district. The results suggest that higher quality hotels are more expensive and price more steeply (in absolute terms) when approaching their capacity constraints. Denote this estimated optimal pricing function as $\hat{p}^*(q)$.

The second component we need in any counterfactual simulation is equilibrium demand. In principle, we can calculate market shares given parameter estimates and some price vector $\boldsymbol{p}$ using Equation (3.2). However, these counterfactual market shares do not necessarily adhere to hotel capacity constraints. We next propose an iterative method to allocate predicted demand to hotels while taking into account their capacity constraints.

For now, let the price vector $\boldsymbol{p}$, the other product characteristics, as well as demand parameters be fixed. Using Equation (3.2), we can then calculate predicted occupied rooms for each product. Denote the vector of predicted occupancy as $\boldsymbol{q_\tau}(\boldsymbol{J_\tau}, \mathbb{d}_\tau)$, where $\boldsymbol{J_\tau}$ denotes the set of products used as choice set in Equation (3.2) in iteration $\tau$ and $\mathbb{d}_\tau$ denotes the number of travelers. We make the dependence of $\boldsymbol{q_\tau}$ on $\boldsymbol{J_\tau}$ and $\mathbb{d}_\tau$ explicit because these are the only components of demand that change in each iteration. Note that the dimension of vector $\boldsymbol{q_\tau}$ equals the number of products in $\boldsymbol{J_\tau}$ and can change with each iteration. For the first iteration $\tau = 1$, let $\boldsymbol{J}_1$ be the set of all products and $\mathbb{d}_1$ be the full market size. In each iteration $\tau$, we proceed as follows:

1. Calculate $\boldsymbol{q_\tau}(\boldsymbol{J_\tau}, \mathbb{d}_\tau)$ using Equation (3.2).

2. Let $q_\tau^j$ denote the jth element of $\boldsymbol{q_\tau(J_\tau, \mathbb{d}_\tau)}$. Calculate $\tilde{q}_\tau^j = \tilde{q}_{\tau-1}^j + q_\tau^j$ for all $j \in \boldsymbol{J_\tau}$. For the first iteration, use $\tilde{q}_0^j = 0$.

3. Let $k^j$ denote the room capacity of product $j$. Denote as $\boldsymbol{J_\tau^c}$ the set of products for which $\tilde{q}_\tau^j > k^j$. Denote the complementary set as $\boldsymbol{J_\tau^{\not c}} = \boldsymbol{J_\tau} \setminus \boldsymbol{J_\tau^c}$. Calculate aggregate excess demand as $\hat{Q}_\tau = \sum_{j \in \boldsymbol{J_\tau^c}} (\tilde{q}_\tau^j - k^j)$.

4. Set final demand $q^j = k^j$ for all $j \in \boldsymbol{J_\tau^c}$. If $\hat{Q}_\tau = 0$ stop, else set $\boldsymbol{J_{\tau+1}} = \boldsymbol{J_\tau^{\not c}}$ and $\mathbb{d}_\tau = \hat{Q}_\tau$ and continue with step 1 of iteration $\tau + 1$.

The basic idea of the iterative process is that in each iteration, predicted hotel demand is allocated to the products up to their capacity constraints. In the next iteration, demand is predicted again but only using the excess demand from the previous iteration and those hotels whose capacities are not yet at their limits. These iterations continue until no excess demand remains.

This algorithm returns a vector of final demand $\boldsymbol{q}$ that contains the demand for each product. Note, however, that so far, the price vector $\boldsymbol{p}$ was kept fixed at some arbitrary values. Given some demand vector $\boldsymbol{q}$, we can now calculate the corresponding optimal hotel prices using $\hat{p}^*(q)$. This yields a vector of optimal prices, say $\boldsymbol{\hat{p}^*(q)}$. The quantities $\boldsymbol{q}$ were, however, obtained based on some other price vector $\boldsymbol{p}$ and are therefore not necessarily consistent with $\boldsymbol{\hat{p}^*(q)}$. Therefore, the iterative process to obtain quantities $\boldsymbol{q}$ described above needs to be nested inside an outer iteration process that determines a price vector that is consistent with the quantities. For each iteration $\kappa$ of this outer loop, we then:

1. Calculate $\boldsymbol{q_\kappa}$ using price vector $\boldsymbol{p_\kappa}$ and the algorithm described above.

2. Calculate $\boldsymbol{\hat{p}^*(q_\kappa)}$.

3. If $\max|\boldsymbol{\hat{p}^*(q_\kappa)} - \boldsymbol{p_\kappa}| \leq \epsilon$ stop, else choose $\boldsymbol{p_{\kappa+1}}$ such that $|\boldsymbol{p_{\kappa+1}} - \boldsymbol{\hat{p}^*(q_\kappa)}| \leq |\boldsymbol{p_\kappa} - \boldsymbol{\hat{p}^*(q_\kappa)}| - \delta$ and continue with step 1 of iteration $\kappa + 1$.

In step 3 of each iteration of the outer loop, we update the price vector incrementally toward $\boldsymbol{\hat{p}^*(q)}$ by an amount $\delta$. The algorithm stops when each element in $\boldsymbol{p_\kappa}$ lies less than a threshold $\epsilon$ away from the optimal price vector $\boldsymbol{\hat{p}^*(q_\kappa)}$. In our simulation, we set $\delta = \epsilon = 10^{-5}$.[38] This double iteration procedure allows us to simulate market equilibria while taking into account the simultaneity of prices and quantities for hotels. In the end, we obtain a price vector $\boldsymbol{p^*}$ and a quantity vector $\boldsymbol{q}$ that are consistent with one another.

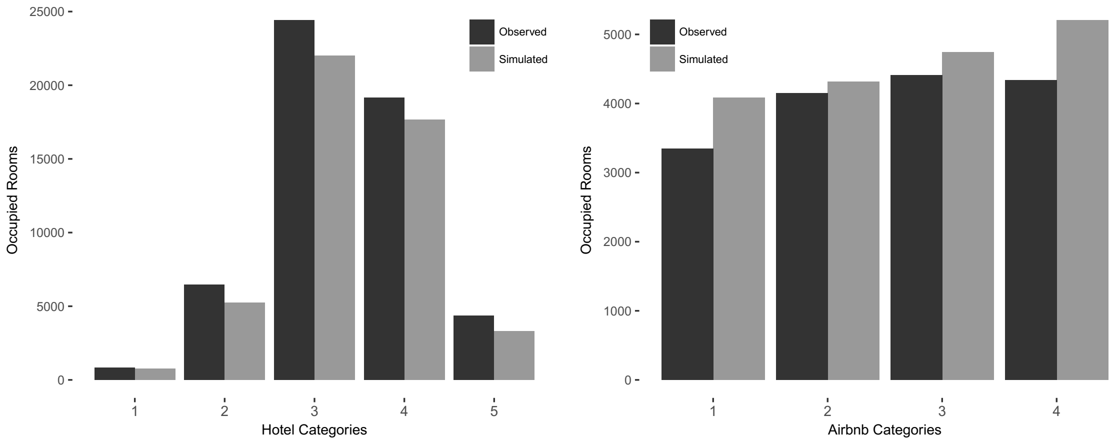
Our supply-side model provides an ad-hoc solution to adjust simulated market shares to comply with hotel capacity constraints. Since we do not estimate hotel costs, we cannot infer anything about hotel profits. Under the assumption of constant markups, changes in profit would be proportional to changes in revenues.

[38]In each iteration step, we could also set $\boldsymbol{p_{\kappa+1}} = \boldsymbol{\hat{p}^*(q_\kappa)}$. However, we find that the algorithm does not converge when proceeding this way. Approaching the optimal price vector incrementally leads to convergence. Varying $\delta$ and $\epsilon$ progressively leads to convergence the fastest: We start by setting $\delta = \epsilon = 1$. After convergence is reached for these initial values of $\delta$ and $\epsilon$, we proceed with the iterative procedure setting $\delta = \epsilon = 0.1$. We continue in this fashion, reducing $\delta$ and $\epsilon$ by a factor of $10^{-1}$, until convergence is reached for $\delta = \epsilon = 10^{-5}$.

### 3.5.3 Model Fit

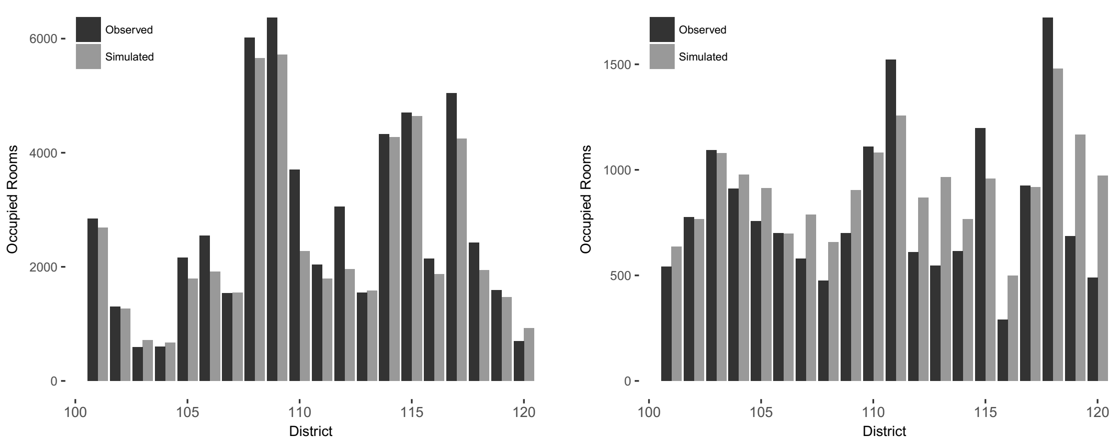
Before analyzing counterfactual scenarios without Airbnb, we assess how well our model matches the observed market outcomes in the benchmark scenario when Airbnb is present in the market. For the simulations, we assume that consumer utility follows the estimates reported in column (5) of Table 3.1 because this specification corrects for changes in Airbnb supply as well as high-hotel-capacity district fixed effects.

Figures 3.10a and 3.10b show the average daily number of occupied rooms by hotel and Airbnb quality categories, respectively. The gray bars show the quantities predicted by our model, the black bars show the quantities that we observe in the data. In general, we tend to overestimate Airbnb demand and to underestimate hotel demand. All in all, our model seems to capture the broader demand patterns reasonably well.




**(a)** Hotels **(b)** Airbnb

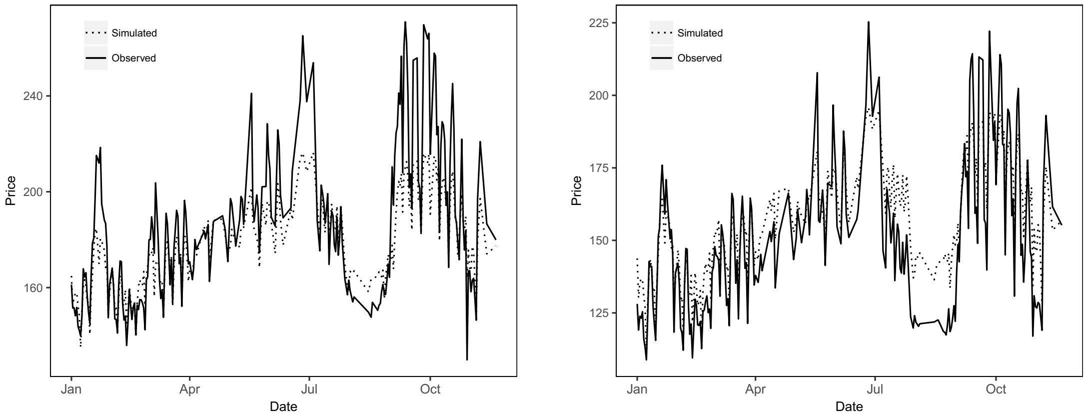
**Figure 3.10:** Simulated vs. observed demand by quality category




**(a)** Hotels **(b)** Airbnb

**Figure 3.11:** Simulated vs. observed demand by districts

Figures 3.11a and 3.11b show the average daily number of occupied rooms by district for hotels and Airbnb, respectively. Generally, we seem to simulate hotel demand well. For Airbnb listings,




**(a)** Including five-star hotels **(b)** Excluding five-star hotels

**Figure 3.12:** Simulated vs. observed hotel prices

the simulation error appears to be more pronounced. Overall, our model captures the geographic distribution of demand well albeit with lower variance.

Turning to simulated hotel prices, our simulation manages to track the sales-weighted average daily prices, as shown in Figure 3.12a. However, in periods of exceptional price peaks, simulated prices tend to be lower than observed prices. This effect appears to be entirely driven by five-star hotels. If we compare the predicted sales-weighted average daily prices with the observed prices for all other hotel categories, our model matches the price data well (see Figure 3.12b).[39]

### 3.5.4 Simulation Results for Counterfactual Scenarios

Based on our hotel supply-side model, we can compare simulated prices and quantities for the benchmark scenario in which Airbnb and hotels are both in the market with prices and quantities when Airbnb is absent. We study two counterfactual scenarios: in the first, hotels are allowed to readjust their prices when Airbnb leaves the market. In the second scenario, hotels are not allowed to adjust their prices. Analyzing both scenarios allows us to disentangle the impact of increased product variety from the impact of hotel price adjustments on consumer surplus. These two scenarios correspond to the two counterfactual scenarios presented in Farronato and Fradkin (2018).

For the benchmark scenario, we use the demand parameters reported in column (5) of Table 3.1. For the counterfactual scenarios, we assume that demand follows a one-stage nested logit model: consumers first choose the district and subsequently choose the option within the district that maximizes utility. The taste correlation between hotels of the same district is taken from our estimate for the taste correlation between accommodations in the same subnest ($\sigma_1$) in the two-level nested logit model.

Table 3.3 shows the simulated price changes for the different hotel types when Airbnb leaves the market. As expected, absent Airbnb, hotels increase their prices. Table 3.3 also displays the

[39]Few large five-star hotels appear to attract large demand despite charging high prices (even compared to other five-star hotels). Our model does not match the observed shares for these “outlier” hotels well. Although the share of these hotels in the overall demand for accommodation is small, this mismatch heavily affects sales-weighted average prices because of the very high prices these hotels charge.

simulated changes in equilibrium quantities for both counterfactual scenarios: when hotels are allowed to change prices ("pr adj") and when hotels leave prices unchanged ("no pr adj"). In both counterfactual scenarios, demand is larger for all hotel categories than in the benchmark case. For most hotel categories, this quantity increase is larger if no price adjustment is allowed. Surprisingly, the reverse is true for the lowest-category hotels. This result is likely due to the different pricing patterns observed for different hotel types. Since one-star hotels increase prices less sharply in response to higher occupancy ratios, more demand is diverted to them when rivals adjust prices in response to Airbnb leaving the market. Combining these hotel price and quantity changes, we find that the average daily revenue loss for hotels due to Airbnb amounts to 1.8 million euro.

**Table 3.3:** Simulated price and occupancy changes for hotels

| Star Rating | Price | Δ Pr | Δ Pr % | Δ Q | | Δ Q % | |
|---|---|---|---|---|---|---|---|
| | | | | pr adj | no pr adj | pr adj | no pr adj |
| 1 | 92.93 | 4.63 | 5.00 | 157.62 | 144.39 | 16.28 | 14.91 |
| 2 | 101.12 | 10.76 | 10.64 | 847.06 | 944.91 | 16.08 | 17.93 |
| 3 | 131.39 | 18.90 | 14.39 | 2269.21 | 3967.71 | 10.30 | 18.01 |
| 4 | 205.92 | 21.06 | 10.23 | 1478.53 | 3176.55 | 8.27 | 17.76 |
| 5 | 470.70 | 21.04 | 4.47 | 169.49 | 582.18 | 5.05 | 17.36 |

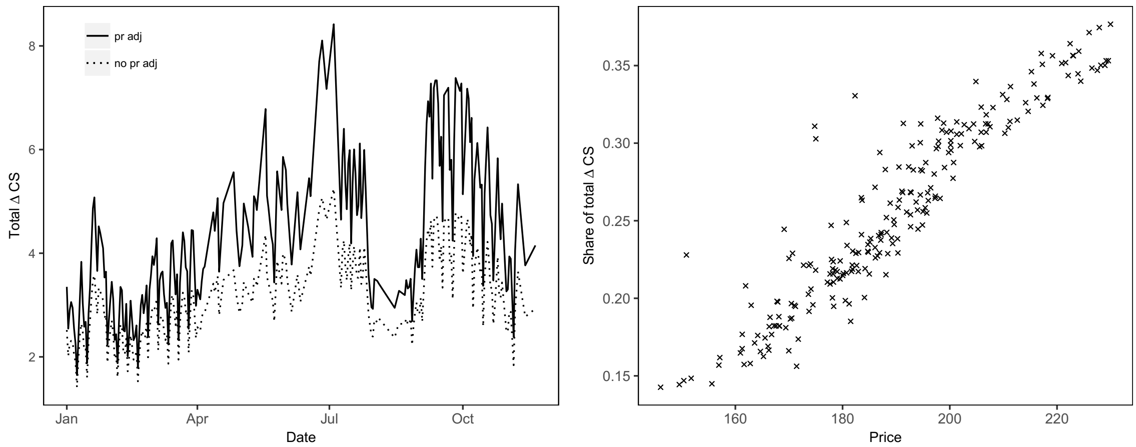
Notes: The second column shows sales-weighted average daily prices when Airbnb is present, the second and third column show resulting price changes when Airbnb is absent. Changes in quantity are calculated for the scenarios with and without price adjustment.

To study the consumer welfare implications of Airbnb, we compute the consumer surplus in each of the simulated scenarios (benchmark, absent Airbnb without price adjustment, absent Airbnb with price adjustment). Then, we compare the consumer surplus in each of the counterfactual scenarios to the benchmark scenario.[40] This comparison allows us to quantify how much consumers gain in terms of consumer surplus in the benchmark scenario with Airbnb compared to the counterfactual scenarios without Airbnb.

When hotels are allowed to adjust prices, we estimate an average gain in consumer surplus from Airbnb of 31 euro per night. In the counterfactual scenario where hotels are not allowed to adjust prices, we estimate the average gain to be only 23 euro. Therefore, lower hotel prices when Airbnb is present account for approximately 26 percent of the gain in consumer surplus from Airbnb. To estimate the total gain in consumer surplus from Airbnb, we multiply the estimated expected gain in consumer surplus per consumer with the estimated market size. We find that, when we allow hotels to adjust prices absent Airbnb, the average daily gain in consumer surplus from Airbnb is estimated to be approximately 4.3 million euro. Ignoring price adjustments, the gain in total consumer surplus is only 3.1 million euro.

Figure 3.13a reveals that the gain in consumer surplus from Airbnb is unevenly distributed across markets. The total consumer surplus gain, accounting for hotel price adjustments, ranges from two to eight million euro. In periods of high demand, when hotels charge high prices, the

[40]The formula for the expected consumer surplus in the nested logit model is an intuitive extension of the well known formula for the logit model and is reported in Appendix 3.7.5.




**(a)** Gain in daily total consumer surplus with and without hotel price adjustment

**(b)** Share in total consumer surplus gain attributable to hotel price adjustment

**Figure 3.13:** Consumer surplus gains from Airbnb with and without hotel price adjustment

total consumer surplus gain from Airbnb is the highest. Figure 3.13b shows that the share of the gain in consumer surplus due to hotel price adjustment is higher when hotels charge higher prices in the benchmark scenario. When prices are high even with Airbnb in the market, hotel price responses can account for an aggregate gain in consumer surplus from Airbnb of up to three million euro. Finally, our model also allows us to assess the share of Airbnb consumers that would choose a hotel option absent Airbnb. We find that, accounting for hotel-price adjustments, only 28 percent of Airbnb consumers would book a hotel option if Airbnb were not available. Without price adjustment, this number amounts to 48 percent.

## 3.6 Conclusion

With the rise of the digital economy, peer-to-peer markets have entered and subsequently disrupted many traditional industries. A prominent example is the rise of Airbnb, from the co-founders hosting their first guests in their apartment in San Francisco in 2007 to featuring more than seven million listings in over 220 countries in 2020.[41] This rapid rise was, however, accompanied by controversial policy debates. One main concern is that Airbnb listings are competing with traditional hotels while subject to much laxer regulation. In this context, it is important to understand the nature of competition between Airbnb and hotels and to assess how and when consumers profit from the online platform.

In this paper, we first investigate the nature of competition between Airbnb listings and hotels in the city of Paris using data from 2017. We combine granular, daily data on hotel occupancy rates, Airbnb demand, and hotel prices to estimate a two-level nested logit model of demand. Our model allows for taste correlation within the different districts of the city as well as within accommodation types (Airbnb listings or hotels) in each district. Our results suggest that the market is segmented both along geography as well as the type of accommodation. More specifically, we find that cross-price elasticities are larger for accommodations located in the same district. Furthermore, within

[41] See `https://news.airbnb.com/fast-facts/` (accessed: July 1, 2020).

each district, cross-price elasticities are even larger for accommodations of the same type, i.e. Airbnb listings or hotel rooms.

Based on these parameter estimates, we next ask what the presence of Airbnb implies for consumer surplus. To address this question, we simulate counterfactual scenarios without Airbnb and compare the outcomes with a scenario comparable to the real world. These simulations allow us to compare consumer surplus, hotel revenue, and consumer choices in a world with and without Airbnb. We find that, on average, consumers gain 31 euro per night in terms of consumer surplus from having Airbnb in the market. Taking into account our estimate of the total market size, this gain amounts to a total average gain in consumer surplus of 4.3 million euro per night. This gain in consumer surplus consists of utility gains because of the increased variety due to Airbnb and lower hotel prices due to increased competition. To disentangle the two, we also regard a counterfactual scenario absent Airbnb in which hotels do not adapt prices. Compared to this scenario, average consumer surplus is 23 euro per night higher in the presence of Airbnb. Therefore, the gain from lower prices due to Airbnb amounts to approximately 26 percent of the entire gain in consumer surplus. Consumers profit most in times when demand is high and hotels approach their capacity constraints. In these situations, the aggregate gain in consumer surplus can amount up to eight million euro per night. In these high demand phases, consumers profit most from being able to choose from Airbnb listings because they would otherwise have to choose between hotels asking high prices due to their high occupancy or not booking an accommodation at all. Concerning hotel revenues, our simulations suggest that they would increase by 1.8 million euro per night absent Airbnb.

Our results are in line with those of Farronato and Fradkin (2018) and Zervas *et al.* (2017). Farronato and Fradkin (2018) find that competition between Airbnb and hotels in the US is particularly fierce when demand is high and hotels are close to capacity-constrained. They document that consumers profit most from Airbnb in these times as hotels charge lower prices than they would absent Airbnb. Zervas *et al.* (2017) find that average hotel revenue decreased in Texas due to the entry of Airbnb. The main mechanism for this result is that hotels charge lower prices due to the increased competition. We complement these results by investigating related aspects on a more fine-grained level. Rather than using data on the city level, we observe and account for variation in demand within the city. Our descriptive results suggest that demand shocks in neighborhoods within the city are an important factor in the demand for short-term accommodation.

All in all, our results suggest that competition between Airbnb and hotels is only moderate most of the times, but Airbnb listings can be an important substitute for consumers in times of high demand. In these times, consumers are able to profit directly from the more flexible supply of Airbnb listings. In general, consumers therefore do seem to profit substantially from the presence of Airbnb in the market.

There is one main caveat to our analysis: We cannot account for geographic proximity of different districts. Therefore, the relationship between products in different districts is identical in the model, irrespective of their geographic proximity. Methods that allow to model geographic relationships more precisely are either much more computationally involved or introduce additional complexity when handling endogeneity (Fox *et al.*, 2016). Therefore, our decision to use the two-level nested logit is motivated by its tractability, as well as ability to incorporate instrumental variables using standard techniques.

Of course, our paper only analyzes one part of the greater scheme when it comes to regulating Airbnb. In particular, because we do not estimate hotel profits, we cannot say anything about how Airbnb impacts those. Furthermore, our analysis abstracts from externalities that Airbnb can create in other markets such as the housing market (Horn and Merante, 2017; Koster *et al.*, 2018; Garcia-López *et al.*, 2020; Barron *et al.*, 2020; Duso *et al.*, 2020) or on local neighborhoods (Basuroy *et al.*, 2020). Therefore, while our results suggest that Airbnb increases consumer welfare for travelers, it is not clear whether this increase is sufficient to offset other potential welfare losses that are relevant when reflecting on total welfare.

## 3.7 Appendix

### 3.7.1 Descriptives

**Table 3.4:** Prices, occupancy, and offer

| Quality | Price | | Rooms Occupied | | Rooms Offered | |
|---|---|---|---|---|---|---|
| | Hotels | Airbnb | Hotels | Airbnb | Hotels | Airbnb |
| 1 | 81.33 | 41.43 | 1112.61 | 4214.60 | 1583.48 | 12577.39 |
| | (15.73) | (1.60) | (186.98) | (847.23) | (11.37) | (1267.82) |
| 2 | 99.63 | 64.03 | 6599.36 | 4258.33 | 8895.86 | 12885.61 |
| | (17.00) | (1.94) | (1228.55) | (849.14) | (86.78) | (1304.51) |
| 3 | 125.93 | 92.67 | 24714.45 | 4550.64 | 31748.35 | 14160.72 |
| | (23.02) | (2.92) | (3906.06) | (964.24) | (131.52) | (1291.13) |
| 4 | 192.81 | 172.83 | 19708.64 | 4539.22 | 25502.31 | 15036.03 |
| | (31.89) | (8.07) | (3093.73) | (1138.67) | (388.45) | (1118.71) |
| 5 | 485.10 | - | 4579.63 | - | 6483.31 | - |
| | (79.59) | - | (796.29) | - | (225.32) | - |

Notes: The table shows averages with standard deviations in parentheses. The prices shown are average prices per night. For hotels, prices are calculated based on dates for which less than 65 percent of prices are missing.

**Table 3.5:** Matched vs. non-matched hotels: Hotel level

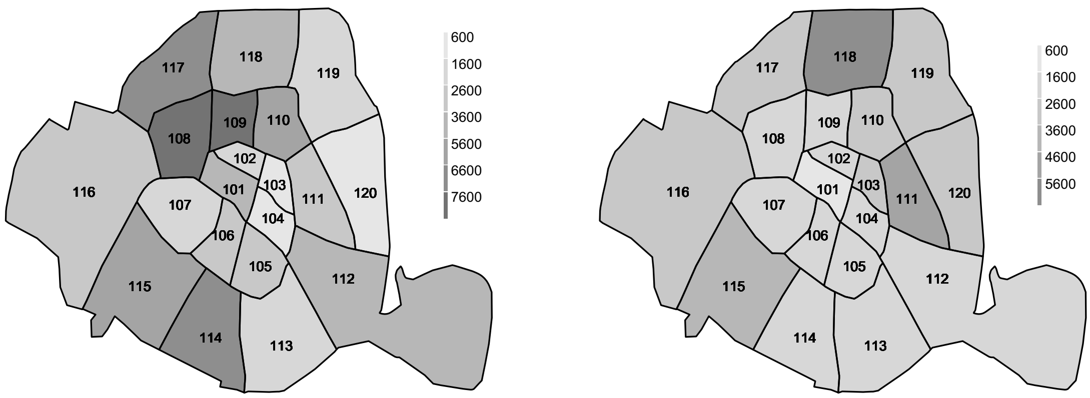
| Star Rating | | Rooms Occupied | | Rooms Offered | |
|---|---|---|---|---|---|
| | Share | Matched | Non-matched | Matched | Non-matched |
| 1 | 33.33 | 45.39 | 21.08 | 63.93 | 31.26 |
| | | (77.89) | (10.52) | (102.47) | (14.53) |
| 2 | 26.75 | 28.69 | 31.39 | 38.77 | 41.73 |
| | | (33.86) | (33.31) | (43.36) | (41.80) |
| 3 | 32.26 | 33.85 | 36.75 | 43.37 | 47.11 |
| | | (32.44) | (39.26) | (40.16) | (52.17) |
| 4 | 45.13 | 59.91 | 51.63 | 78.10 | 65.75 |
| | | (96.16) | (68.56) | (125.52) | (83.51) |
| 5 | 32.89 | 58.18 | 74.35 | 82.95 | 103.65 |
| | | (39.51) | (66.74) | (53.78) | (95.2) |

Notes: The second column shows the percentage of hotels that we could not match for each category. The other statistics are the average daily number of occupied and offered rooms per hotel, calculated over the entire year (standard deviations in parentheses). The unit of observation is a single hotel.

**Table 3.6:** Matched vs. non-matched hotels: Room level

| Star Rating | | Rooms Occupied | | Rooms Offered | |
|---|---|---|---|---|---|
| | Share | Matched | Non-matched | Matched | Non-matched |
| 1 | 19.25 | 907.73 | 205.50 | 1278.78 | 304.71 |
| | | (153.63) | (34.94) | (8.94) | (9.13) |
| 2 | 27.79 | 4753.02 | 1859.66 | 6423.65 | 2472.21 |
| | | (895.84) | (350.75) | (49.79) | (46.05) |
| 3 | 35.84 | 15900.61 | 8877.17 | 20370.08 | 11378.27 |
| | | (2579.40) | (1397.01) | (58.68) | (131.76) |
| 4 | 40.25 | 11686.96 | 8061.20 | 15236.77 | 10265.54 |
| | | (1881.30) | (1305.69) | (522.86) | (387.19) |
| 5 | 35.48 | 2934.05 | 1650.19 | 4182.78 | 2300.53 |
| | | (488.89) | (321.76) | (60.70) | (171.71) |

Notes: This table is similar to Table 3.5. However, rather than comparing hotel-level averages, we compare aggregate room capacities and occupation by hotel category. The second columns shows the percentage of non-matched rooms for each hotel category. The other statistics are the average daily number of occupied and offered rooms per hotel category, calculated over the entire year (standard deviations in parentheses). This table shows that we capture the majority of hotel rooms in the sample of matched hotels.




**(a)** Hotels **(b)** Airbnb

**Figure 3.14:** Average geographical distribution of hotel rooms and Airbnb listings over the year. The legends show the lower limit of brackets, i.e. "600" represents districts with Airbnb or hotel numbers between 600 and 1,600 rooms. The scales in both figures were chosen to allow for comparison between both accommodation types. The minimum hotel capacity in a district is 608, the maximum 8524. The minimum Airbnb offer in a district is 1338, the maximum 6130. In specification (5) of Table 3.1, we include hotel-specific fixed effects for the following districts: 101, 108, 109, 114, 115 and 117.

### 3.7.2 Market Size as Instrument

In this appendix, we show by simulation that market size can act as a valid instrument. We have four single-product firms. For each firm $j$, the econometrician observes two characteristics $x_j$ and $p_{jt}$. In each market $t$, we have $N_t$ individuals that either choose one of the four products or the outside good $j = 0$ with utility normalized to zero. Note that the market size varies by market. For 500 markets, we simulate the following scenario:

1 We set the vector of products characteristics $x$ equal to $(0.4, 0.5, 0.9, 1)$ and the price vector $p$ equals to $(0.2, 0.3, 0.4, 0.5)$.

2 Utility of consumers is given by $u_{ij} = x_j - 2 \times p_{jt} + \mathcal{E}_{jt} + \epsilon_{jt}$, where $\mathcal{E}_{jt}$ are unobserved product- and market-specific shocks and $\epsilon_{jt}$ is i.i.d extreme value type 1.

3 For each market $t$, the market size $N_t$ is randomly drawn from the vector (500, 600, 700, 800, 1000).

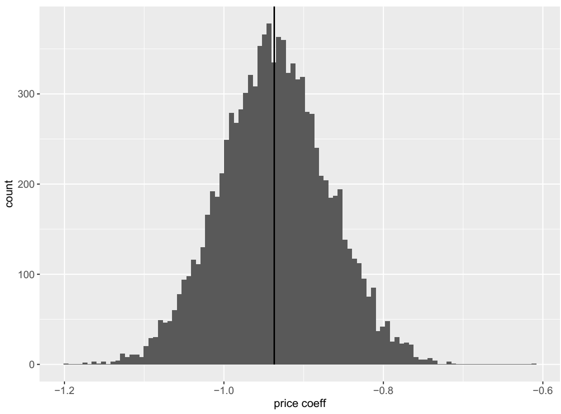
4 Firms observe the market size $N_t$ and set prices according to $p_{jt} = p_j + 0.001 \times N_t$

5 Additionally in each period *one* product $j$ is randomly shocked with $\mathcal{E}_{jt} \sim U[0, 1]$. If product $j$ is shocked, the firm sets the price $p_{jt} = p_j + 0.001 \times N_t + 0.1 \times \mathcal{E}_{jt}$

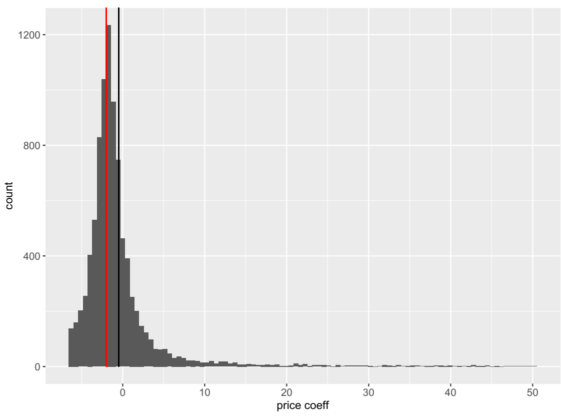
The above situation emulates a scenario where firms respond to changes in the market size by uniformly raising prices. Additionally, in each period, one firm experiences a demand shock that it observes and raises prices accordingly. This emulates the idea of localized demand shocks. We simulate the aggregate market outcomes and perform three different estimations: (a) a simple logit, (b) an IV-logit with classical BLP instruments, i.e. the mean over rivals' $x_{jt}$, and (c) an IV-logit with the BLP instruments interacted with the market size $N_t$.

Note that the classical BLP instrument is collinear with the vector $x$ in this setting. The reason is that the classical BLP instruments only work when either the choice set or the characteristics of the products vary by market. Therefore, in our simulation, we randomly drop one product in each market situation. When interacting the BLP instrument with the market size, this collinearity does not arise.

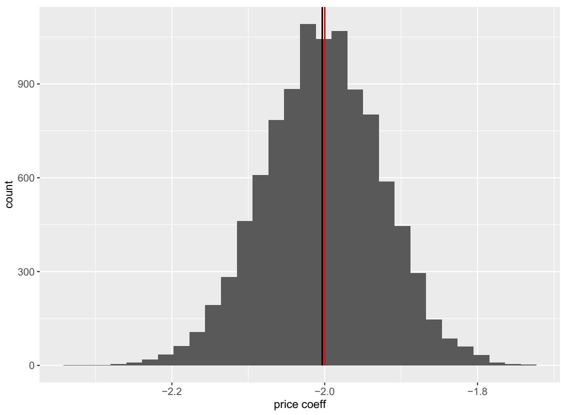
We perform 10,000 Monte Carlo runs. In each run, we simulate the aggregate choices and estimate the structural parameters. Figure 3.15a shows the histogram for the estimated price coefficient in each of the simulation runs for the logit model. Recall that the true price coefficient is -2. As expected, in the naive model, we underestimate the magnitude of the true price parameter. The bias is approximately equal to 1.4, on average. Figure 3.15b shows the result for the classical BLP instrumental variable technique. The performance of the estimator is surprisingly poor. In some simulation runs, we obtain extremely large deviations from the true parameter value. Figure 3.15b therefore only shows the distribution of values while truncating outliers. The mean over the Monte Carlo runs is equal to 3.99 with a variance of 218,443. The median is equal to -1.81. By contrast, the estimator that interacts the classical BLP instruments with the market size is well-behaved and appears to be consistent with a mean of -2.003, a variance of 0.006 and a median of -2.002 (see Figure 3.15c).




**(a)** Logit




**(b)** IV-logit “BLP”




**(c)** IV-logit “BLP $\times$ $N_t$”

**Figure 3.15:** Histograms of estimated price coefficients. The red line shows the true price coefficient, the black line shows the mean estimated price coefficients.

### 3.7.3 First-stage F-statistics

Table 3.7 reports the F-statistics for the excluded instruments for the various specifications reported in Table 3.1. We report the F-statistic obtained for each endogenous variable in each specification. Note that the F-statistic for price in the specification of column (2) of Table 3.1 is identical to the F-statistic for the price in the specification of column (3) of Table 3.1.

**Table 3.7:** F-statistics corresponding to Table 3.1

| | (2)/(3) Logit/NL | (4) NL | (5) NL |
|---|---|---|---|
| Price | 602.83 | 648.81 | 3631.20 |
| $\sigma_1$ | 257.26 | 263.96 | 4906.98 |
| $\sigma_2$ | 7636.45 | 8045.26 | 11793.30 |

### 3.7.4 Robustness Checks

**Table 3.8:** Results with alternative market size definition

| | (1) Logit | (2) Logit (IV) | (3) NL | (4) NL | (5) NL |
|---|---|---|---|---|---|
| Price | 0.002*** | −0.008*** | −0.006*** | −0.007*** | −0.005*** |
| | (0.0001) | (0.0004) | (0.0003) | (0.0003) | (0.0003) |
| $\sigma_1$ | | | 0.770*** | 0.645*** | 0.712*** |
| | | | (0.018) | (0.014) | (0.018) |
| $\sigma_2$ | | | 0.234*** | 0.228*** | 0.207*** |
| | | | (0.010) | (0.009) | (0.006) |
| Airbnb category 2 | 0.289*** | 0.506*** | 0.065*** | 0.140*** | −0.015 |
| | (0.041) | (0.045) | (0.022) | (0.023) | (0.011) |
| Airbnb category 3 | 0.420*** | 0.935*** | 0.278*** | 0.394*** | 0.084*** |
| | (0.040) | (0.047) | (0.028) | (0.028) | (0.018) |
| Airbnb category 4 | 0.259*** | 1.617*** | 0.782*** | 0.942*** | 0.469*** |
| | (0.042) | (0.068) | (0.045) | (0.045) | (0.039) |
| 1-star hotel | −0.544*** | −0.144*** | 0.742*** | 0.602*** | 2.576*** |
| | (0.048) | (0.053) | (0.035) | (0.031) | (0.066) |
| 2-star hotel | 0.739*** | 1.333*** | 1.000*** | 1.037*** | 2.744*** |
| | (0.041) | (0.049) | (0.025) | (0.026) | (0.094) |
| 3-star hotel | 2.065*** | 2.942*** | 1.483*** | 1.699*** | 3.275*** |
| | (0.041) | (0.055) | (0.042) | (0.041) | (0.121) |
| 4-star hotel | 1.641*** | 3.217*** | 1.854*** | 2.073*** | 3.553*** |
| | (0.043) | (0.076) | (0.052) | (0.051) | (0.127) |
| 5-star hotel | −0.220*** | 4.427*** | 3.287*** | 3.526*** | 4.537*** |
| | (0.067) | (0.180) | (0.117) | (0.113) | (0.167) |
| Log Airbnb offer | | | | | 0.356*** |
| | | | | | (0.016) |
| Constant | −6.486*** | −6.036*** | −4.174*** | −4.842*** | −6.866*** |
| | (0.039) | (0.044) | (0.046) | (0.071) | (0.128) |
| Market FEs | *N* | *N* | *N* | *Y* | *Y* |
| N | 33,504 | 33,504 | 33,504 | 33,504 | 33,504 |
| Adjusted $R^2$ | 0.374 | 0.120 | 0.641 | 0.697 | 0.861 |

Notes: These results correspond to the main analysis but using the instantaneous market size measure obtained from the Google Search data. The lowest Airbnb quality category serves as the base category. *, **, *** indicate statistical significance at the ten, five, and one percent level, respectively.

**Table 3.9:** Results using period from February 1st to July 1st

| | (1) Logit | (2) Logit (IV) | (3) NL | (4) NL | (5) NL |
|---|---|---|---|---|---|
| Price | 0.002*** | −0.009*** | −0.007*** | −0.007*** | −0.005*** |
| | (0.0002) | (0.001) | (0.0005) | (0.0004) | (0.0004) |
| $\sigma_1$ | | | 0.703*** | 0.666*** | 0.786*** |
| | | | (0.026) | (0.020) | (0.027) |
| $\sigma_2$ | | | 0.224*** | 0.220*** | 0.211*** |
| | | | (0.016) | (0.013) | (0.009) |
| Airbnb category 2 | 0.317*** | 0.556*** | 0.128*** | 0.147*** | −0.025 |
| | (0.062) | (0.069) | (0.036) | (0.034) | (0.018) |
| Airbnb category 3 | 0.472*** | 1.033*** | 0.387*** | 0.412*** | 0.081*** |
| | (0.061) | (0.073) | (0.046) | (0.042) | (0.028) |
| Airbnb category 4 | 0.311*** | 1.783*** | 0.966*** | 0.985*** | 0.474*** |
| | (0.064) | (0.107) | (0.074) | (0.066) | (0.055) |
| 1-star hotel | −0.513*** | −0.079 | 0.725*** | 0.716*** | 2.369*** |
| | (0.072) | (0.080) | (0.055) | (0.045) | (0.101) |
| 2-star hotel | 0.907*** | 1.553*** | 1.158*** | 1.167*** | 2.444*** |
| | (0.062) | (0.075) | (0.042) | (0.039) | (0.145) |
| 3-star hotel | 2.213*** | 3.169*** | 1.755*** | 1.803*** | 2.870*** |
| | (0.062) | (0.086) | (0.069) | (0.061) | (0.185) |
| 4-star hotel | 1.778*** | 3.464*** | 2.146*** | 2.181*** | 3.156*** |
| | (0.065) | (0.120) | (0.085) | (0.075) | (0.191) |
| 5-star hotel | −0.104 | 4.782*** | 3.674*** | 3.657*** | 4.145*** |
| | (0.098) | (0.278) | (0.185) | (0.163) | (0.237) |
| Log Airbnb offer | | | | | 0.288*** |
| | | | | | (0.024) |
| Constant | −6.667*** | −6.179*** | −4.423*** | −4.685*** | −6.184*** |
| | (0.060) | (0.069) | (0.074) | (0.087) | (0.191) |
| Market FEs | *N* | *N* | *N* | *Y* | *Y* |
| N | 14,773 | 14,773 | 14,773 | 14,773 | 14,773 |
| Adjusted $R^2$ | 0.401 | 0.160 | 0.638 | 0.721 | 0.879 |

Notes: These results correpond to the main analysis, but using only the sample from February to July 2017. Figure 3.1 shows that during this period, the number of Airbnb listings varies less than in the second half of the year. The lowest Airbnb quality category serves as the base category. *, **, *** indicate statistical significance at the ten, five, and one percent level, respectively.

## 3.7.5 Formulas for Elasticities and Consumer Surplus

In this section, we present the formulas for the own- and cross-price elasticities in the two-level nested logit model and the consumer surplus in logit and nested logit models.

### 3.7.5.1 Elasticities

We start with the formulas of the cross-price elasticities. We follow the notation used by Verboven (1996):

$$-e_{jj} = \frac{\partial q_j}{\partial p_j}\frac{p_j}{q_j} = -\alpha p_j\Big(\frac{1}{1-\sigma_1} - \big(\frac{1}{1-\sigma_1} - \frac{1}{1-\sigma_2}\big)\frac{s_j}{s_{hd}} - \frac{\sigma_2}{1-\sigma_2}\frac{s_j}{s_d} - s_j\Big) \tag{3.7}$$

$$e_{jk} = \frac{\partial q_k}{\partial p_j}\frac{p_j}{q_k} = \alpha p_j\Big(\big(\frac{1}{1-\sigma_1} - \frac{1}{1-\sigma_2}\big)\frac{s_j}{s_{hd}} + \frac{\sigma_2}{1-\sigma_2}\frac{s_j}{s_d} + s_j\Big) \tag{3.8}$$

$$e_{jk'} = \frac{\partial q_{k'}}{\partial p_j}\frac{p_j}{q_{k'}} = \alpha p_j\Big(\frac{\sigma_2}{1-\sigma_2}\frac{s_j}{s_d} + s_j\Big) \tag{3.9}$$

$$e_{jk''} = \frac{\partial q_{k''}}{\partial p_j}\frac{p_j}{q_{k''}} = \alpha p_j s_j \tag{3.10}$$

### 3.7.5.2 Consumer Surplus

In a two-level nested logit model, the formula for the expected utility of consumers is given by:

$$CS = ln\Big(\sum_{d=0}^{D} e^{I_{dt}}\Big) \tag{3.11}$$

Where $I_{dt}$ is defined by:

$$I_{dt} = (1-\sigma_2)ln\Big(\sum_{h\in V_{dt}} e^{I_{hdt}/(1-\sigma_2)}\Big) \tag{3.12}$$

and $I_{hdt}$ is defined by:

$$I_{hdt} = (1-\sigma_1)ln\Big(\sum_{l\in V_{hdt}} e^{\delta_{jt}/(1-\sigma_1)}\Big) \tag{3.13}$$

If we multiply Equation (3.11) by the negative inverse of the estimated price coefficient, we obtain the euro value of the consumer surplus. The formula for the consumer surplus is a natural extension of the consumer surplus for the simple logit. Equation (3.11) is the logsum over the denominators of the nests (which are multiplied with the nest parameters). Each denominator of a nest is itself a logsum over the denominators of the subnests contained in the nest (multiplied with the subnest parameters).

## 3.7.6 Semi-Elasticities

**Table 3.10:** Average estimated demand semi-elasticities ×100

| | | | Cross-price elasticities | | |
|---|---|---|---|---|---|
| Type | Category | Own-price elasticity | Same district & Type | Same district | Other district |
| Airbnb | 1 | -1.3217 | 0.4591 | 0.0161 | 0.0013 |
| | 2 | -1.5037 | 0.2771 | 0.0093 | 0.0007 |
| | 3 | -1.4810 | 0.2998 | 0.0098 | 0.0008 |
| | 4 | -1.4058 | 0.3750 | 0.0111 | 0.0009 |
| Hotels | 1 | -1.6924 | 0.0884 | 0.0071 | 0.0007 |
| | 2 | -1.6476 | 0.1332 | 0.0110 | 0.0011 |
| | 3 | -1.2696 | 0.5112 | 0.0431 | 0.0042 |
| | 4 | -1.3463 | 0.4345 | 0.0360 | 0.0033 |
| | 5 | -1.3986 | 0.3822 | 0.0353 | 0.0039 |

# Chapter 4

# Airbnb and Rents: Evidence from Berlin[1]

**Bibliographic information: Duso, T., Michelsen, C., Schaefer, M., Tran, K.D., 2020. Airbnb and Rents: Evidence from Berlin, DIW Discussion Paper No. 1890.**

## 4.1 Introduction

Cities worldwide are regulating online short-term rental platforms, such as Airbnb. One rationale behind these regulations is that short-term renting removes apartments from the long-term rental market, thereby contributing to rising rents. Documenting and quantifying such a causal relationship is, however, challenging. Understanding the effectiveness of short-term rental regulation and whether it actually affects rents is important for assessing these regulations *ex-post* and to guide policy-making with regard to online short-term rental platforms in the future.

In this paper, we address two main issues: First, we assess whether policy changes induced by short-term rental regulation in Berlin (the so-called Zweckentfremdungsverbot-Gesetz (ZwVbG)) affected the supply of apartments on Airbnb, the largest online short-term rental platform. Second, we use the plausibly exogenous variation in Airbnb listings due to the policy changes to analyze the causal link between Airbnb and rents. Furthermore, we shed light on the heterogeneity of the effect with regard both to the geography as well as to characteristics of Airbnb listings and rental apartments.

Our paper contributes to a small, but growing, literature on the causal impact of Airbnb on the housing market. Common approaches to address the inherent endogeneity in establishing this causal link are controlling directly for omitted variables bias and reverse causality (Horn and Merante, 2017), using shift-share-like instruments for Airbnb popularity (Barron *et al.*, 2020; Garcia-López *et al.*, 2020), and using policy changes as natural experiments (Koster *et al.*, 2018). We contribute new evidence based on two policy changes in Berlin and by using instrumental

[1]This chapter is joint work with, Tomaso Duso, Claus Michelsen and Kevin Tran. We thank Empirica AG for giving us access to the rental data. This analysis would not have been possible without the people behind the Inside Airbnb project thanks to whom we gained access to the Airbnb data. We further thank OpenStreetMap and the city of Berlin for providing free access to their data.

variables (IV) methods to establish causality. In addition to our natural experiment setting, we use a wide range of potential covariates in an attempt to exclude omitted variable bias to the fullest possible extent. To deal with this large number of covariates, we use the Lasso-based regression methods proposed by Belloni *et al.* (2014) and Chernozhukov *et al.* (2015) for model selection. The broad set of observable characteristics on rentals and Airbnb listings also allows us to explore the heterogeneity of the effect of Airbnb on rents in more detail. So far, the literature mostly focuses on average effects.

Most of the literature analyzes the topic for different regions in the US (Horn and Merante, 2017; Koster *et al.*, 2018; Barron *et al.*, 2020; Valentin, 2020). Evidence for Europe is scarcer with Garcia-López *et al.* (2020) analyzing the case of Barcelona. Thus, we add empirical evidence for the European context by studying Airbnb in Berlin which, as of early 2019, had the fifth-largest number of Airbnb listings among European cities.[2]

In our analysis, we first document that the two policy reforms aimed at curbing short-term accommodations in the city decreased the average number of entire homes listed on Airbnb by eight to ten listings per square kilometer. We further show that in particular the first policy reform which essentially restricted the misuse of regular apartments as short-term rental accommodations, also strongly decreased the average number of days per year that Airbnb listings are available for booking. Using these policy-induced changes to Airbnb in the city, in a second step, we find that each additional nearby entire home on Airbnb increases monthly rents by at least seven cents per square meter. However, we show that this effect is mostly driven by those Airbnb listings that are available for booking for larger parts of the year. Focusing the analysis on just these high-availability listings, we find an average effect of up to 13 cents per square meter per additional nearby Airbnb listing. This effect of high-availability listings is larger than most results found in the literature. Further, our analysis suggests heterogeneous effects across different districts in the city. The marginal effect of Airbnb listings on rents appears to be greater in districts with a lower Airbnb density. Additionally, our results suggest that the districts with a higher Airbnb density likely experienced a larger slowdown in rent increases due to the policy.

For our analysis, we use web scraped data of listed rentals in Berlin. We combine these data with web scraped data on Airbnb in the city. We further add data on various neighborhood characteristics such as points-of-interest and pollution.

In a first step, we analyze the direct impacts of the ZwVbG in Berlin. The law took effect in May 2014 but included a transition period that ended in May 2016. Toward the end of this transition period (from here on, the "May 2016 reform"), we find an average decrease of approximately eight entire homes on Airbnb per square kilometer in Berlin. The law also received a major update that took effect in August 2018 (from here on, the "August 2018 reform"). This update requires hosts on Airbnb and similar platforms to show a registration number on their listings that can only be acquired from local authorities. As a consequence of this policy change, another large decrease in the number of Airbnb listings in the city is observed. This additional decrease amounts to approximately ten entire homes per square kilometer on average and is similar in size to the first drop toward the end of the transition period. Both policy changes affected entire homes on Airbnb more strongly than other types of Airbnb listings. While the average decrease in the number of entire homes listed on Airbnb appears similar for each of these policy changes, they affected different groups of Airbnb listings. While the May 2016 reform led listings that were available for

[2]The top four European cities in descending order were London, Paris, Rome, and Copenhagen. Compare `https://www.statista.com/statistics/815145/airbnb-listings-in-europe-by-city/` (accessed: July 3, 2020).

booking for large parts of the year to leave the platform, the August 2018 reform mostly caused low-availability listings to drop from Airbnb.

In a second step, we then use these policy changes as instruments to assess the impact of Airbnb on asked rents in the city. Our first step analysis suggests that the reforms are relevant instruments, at least in short windows around the changes. Therefore, we focus on seven-month windows around the policy changes in our main analyses. To deal with the large number of potential covariates and model selection, we employ double Lasso methods suggested by Belloni *et al.* (2014) and Chernozhukov *et al.* (2015). To quantify Airbnb presence in this analysis, we count the number of entire homes listed on Airbnb within 250 meters of each rental apartment. We find that Airbnb indeed causes rent increases. Our base specifications suggest rent increases of seven to ten cents per square meter per additional nearby entire home on Airbnb. However, when we focus on only those Airbnb listings that are available for booking for at least 180 days in a year, the estimated marginal effect lies at ten to 13 cents per square meter.

In a third step, we repeat the main analysis but stratify the sample by districts. Our results suggest substantial effect heterogeneity with effect sizes of up to 46 cents per square meter for each additional nearby Airbnb. We find suggestive evidence that those districts with a lower Airbnb density experience larger marginal rent increases per Airbnb on average. In an additional exercise, we then calculate what our first- and second-stage estimates imply for the cumulative effect of the law on rents in the different districts of the city. While we find lower marginal effects of Airbnb in districts with a higher Airbnb density, these same districts experienced a larger decrease in the number of Airbnb listings. Combining the first- and second-stage results of our heterogeneity analysis, we find that districts with a higher Airbnb density likely experienced a larger slowdown in rent increases due to the policy.

Our research contributes most directly to the literature trying to assess the causality between Airbnb and the housing market. Horn and Merante (2017) assess the impact of Airbnb on listed rents in Boston. They find that a one standard deviation increase in Airbnb listings increases rents by an average of 0.4 percent. The effect is particularly strong for apartments with two or more rooms. Further, they find evidence suggesting that this effect is driven by a decrease in rental supply. To address endogeneity, their analysis is based on observable variables and lagged Airbnb numbers to address omitted variable bias and/or reverse causality. Barron *et al.* (2020) analyze the impact of Airbnb on rents and house prices in the US. Using a shift-share instrumental variable approach, they find that a one percent increase in Airbnb listings increases rents by 0.018 percent and house prices by 0.026 percent on average. Garcia-López *et al.* (2020) use a panel fixed-effects specification as well as a shift-share instrumental variable approach similar to Barron *et al.* (2020) and find that rents and house prices in Barcelona increased as Airbnb listings became more widespread. Koster *et al.* (2018) and Valentin (2020) are the only papers that we are aware of that use policy changes as natural experiments for identification. Koster *et al.* (2018) analyze the impact of Airbnb on house prices and rents in Los Angeles County. They use the fact that several cities in the county started regulating short-term accommodations at different points in time to estimate an RDD model around city borders. They show that the regulation reduced Airbnb listings by approximately 50 percent. In their main results, they further show that this regulation reduces property prices by three percent. In additional estimations at the ZIP-code level, they also find that rents decreased by three percent on average. Valentin (2020) analyzes the introduction of short-term rental regulation in New Orleans, showing that a ban on short-term rentals decreased house prices by 30 percent.

We proceed as follows. In Section 4.2, we present institutional details about the regulation of

Airbnb in Berlin. Further, we describe the different data sets used in the analysis. In Section 4.3, we provide first descriptive analyses. In Section 4.4, we discuss our identification strategy and other methodological issues. In Section 4.5, we show our main results. In Section 4.6, we perform robustness checks. In Section 4.7, we examine the heterogeneity of our results. Section 4.8 concludes.

## 4.2 Institutional Details and Data

Airbnb is a peer-to-peer accommodation platform that was founded in 2008. As of June 2020, hosts offer accommodations in more than 220 countries[3]. Airbnb hosts can decide to offer a shared room, a private room, or a full accommodation for rent. Since September 2019, Airbnb also features hotel rooms.

Many cities around the world have introduced laws to regulate Airbnb over the years. In Berlin, regulation was introduced with the ZwVbG, which was passed on December 12, 2013. This original law only introduces the option to declare a ban on the misuse of apartments for purposes other than long-term rental in areas in Berlin where a sufficient supply of housing is at risk. The law defines several cases that would constitute misuse, among which there are short-term renting, commercial use, and long-term vacancy. For our analysis, the most relevant misuse case is short-term renting. If such a ban is declared, misuse of apartments would only be allowed with permission from local district authorities. In the original version of the law, it was not clearly defined which misuse cases would still be allowed.

On March 4, 2014, the Berlin Senate passed a decree to implement the ZwVbG in Berlin. The decree declared that the supply of housing is at risk across the entire city of Berlin and, therefore, a ban on the misuse of apartments took effect in May 2014. However, the ZwVbG included a two-year transition period during which cases that would be defined as misuse under the law but that were already active before May 2014 would still be permitted. This transition period ended in May 2016. Thereafter, short-term renting would only be allowed with a permit issued by the district in which the apartment is located.

On April 20, 2018, an updated version of the law was passed. This update took effect on August 1, 2018. A major change included was that hosts on Airbnb and similar platforms are now required to display a registration number. In order to obtain this registration number, permission to sublet the apartment has to be obtained from the district. At the same time, the update more clearly defines cases for which permission for short-term renting is to be granted. In particular, residents who would like to rent out their main residence during their own absence are now allowed to do so, as long as the status of the apartment as their main residence is not affected. Furthermore, residents can now also permanently rent out parts of their apartments if these parts make up less than 50 percent of the living space. Secondary residences now qualify for a permit if they are used as short-term accommodation for no more than 90 days per year. Figure 4.1 summarizes the different stages of the law.

For our analysis, we focus on the end of the transition period in May 2016 as well as on the update of the law in August 2018. We refer to these policy changes as the "May 2016 reform" and the "August 2018 reform."

For our analysis, we combine data from multiple sources. To measure rents, we use data pro-

[3]See `https://press.atairbnb.com/about-us/` (accessed June 8, 2020).

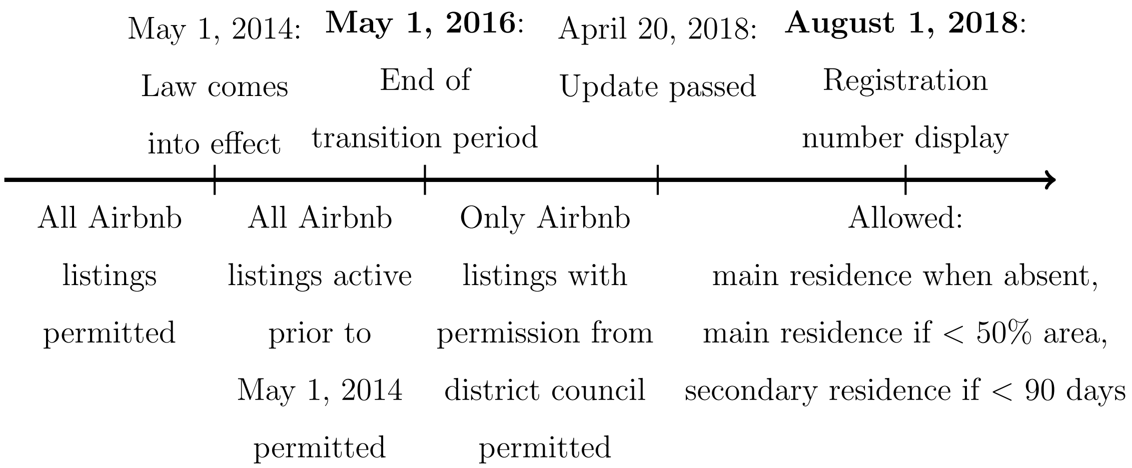



**Figure 4.1:** Stages of the ZwVbG. Dates in bold denote the policy dates used in our analysis.

vided to us by Empirica, an economic consultancy.[4] The data include web scraped information of asked rents for apartments listed in Berlin. The full data set features all rental apartments listed from January 2013 through July 2019. Besides rents, the data also contain various apartment characteristics, such as the size of the apartment and the number of rooms, all of which might be important for explaining rents.

To assess the number of Airbnb listings in Berlin over time, we use data that is publicly available on the website InsideAirbnb as well as data provided by Tom Slee. Both sources provide monthly snapshots of web scraped data of Airbnb listings for multiple cities in the world. Data for Berlin are available in monthly intervals from May 2014 to May 2020, however with several gaps inbetween.

To account for heterogeneity in the attractiveness of different neighborhoods, we use geographical data from OpenStreetMap. The data include geographic information for various types of points-of-interest such as bus stops, restaurants, and supermarkets.[5] Note that the data we use are a snapshot as of February 2018 and, thusly, offer only cross-sectional variation. Further, we use data provided by the city of Berlin that include various variables at a local level such as the amount of noise at night or the level of particulate matter in the air.[6] Again, these data are only of cross-sectional nature.

Our analysis uses data aggregated at two different levels of aggregation. In our main analysis, we use data on a disaggregated level. Each observation constitutes one rental in the month in which it was first listed for rent. This granular perspective has several advantages. First, it allows us to directly use the various apartment characteristics that are included in our rent data to account for aspects that might impact rents. Second, we can use information on the distribution of Airbnb listings in the city in a more detailed manner. More specifically, to measure Airbnb exposure, we count all entire homes listed on Airbnb in a circle of 250 meter radius surrounding each apartment.[7]

[4] See `https://www.empirica-institut.de/en/company-profile/`.

[5] The data are available at `https://download.geofabrik.de/europe/germany/berlin.html`.

[6] The data are available at `https://fbinter.stadt-berlin.de/fb/index.jsp`.

[7] The choice of a radius of 250m is ad-hoc. A trade-off is involved: On the one hand, we expect Airbnb listings that are closer to an apartment to have a larger impact on rents. On the other hand, the location of each Airbnb listing is shifted by up to 150m (see `http://insideairbnb.com/about.html` (accessed: June 8, 2020)). Larger circle sizes would therefore decrease the importance

For the main analysis, we focus on entire homes listed on Airbnb because we expect these to have the largest effect on rents.[8] Figure 4.2 illustrates the calculation of our measure of Airbnb exposure.

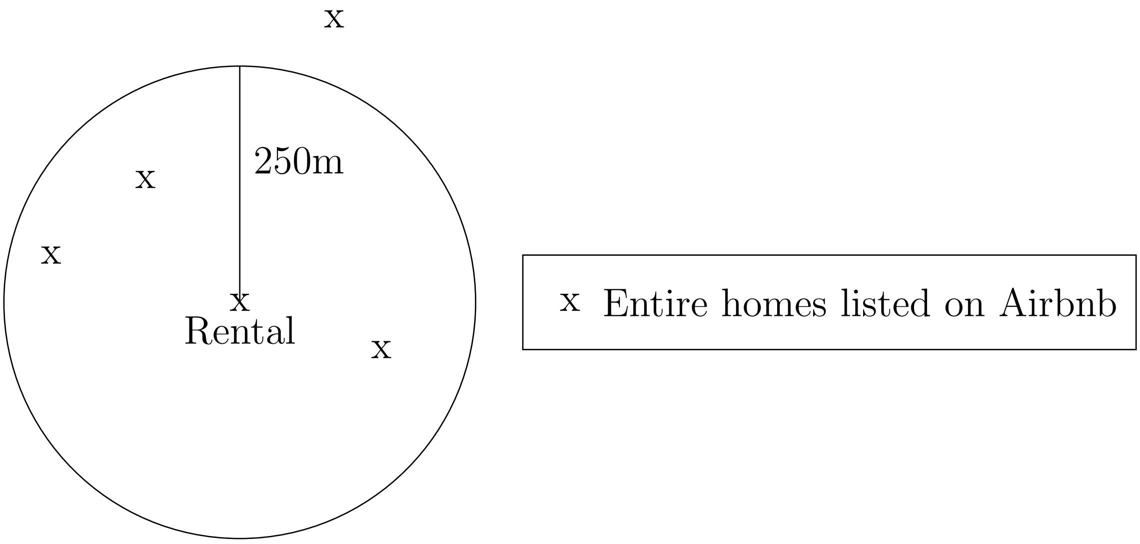



**Figure 4.2:** Illustration of our measure for Airbnb exposure. In this example, the listing would be assigned to have three Airbnb listings nearby.

In additional analyses, we aggregate the disaggregated data set to geographical areas called Lebensweltlich orientierte Räume (LOR). LORs are statistical areas defined and used by the city of Berlin. They are supposed to capture areas that are similar in livelihood. As of January 2019, the city of Berlin is divided into 448 different LORs.[9] This aggregated perspective allows us to conduct additional analyses, such as assessing the impact of the policy changes on the number of rentals listed in the city.

Because we want to use detailed geographical information in our main analysis, we restrict our sample to those rentals for which exact address information is available. This is the case for approximately 80 percent of the data set. Because the data do not include coordinates, we use the address information to geocode the location of each rental. This could be done without any issues for approximately 95 percent of the rentals for which we have the full address information. We exclude the rest. This leaves us with approximately 76 percent of the original observations (212,831 observations for the full sample).

Our main analyses focus on the policy changes of May 2016 and August 2018. We conduct our main analyses for seven-month time windows around these two policy changes (three pre-treatment months, three post-treatment months, and the treatment month).

of the resulting measurement error. Using larger Airbnb circle sizes reduces the effect sizes that we find, in line with our intuition. We report these results in Section 4.6.2.

[8] If we use private or shared rooms, our main estimates are qualitatively similar, but noisier. One reason for this imprecision is that the policy reforms did not affect these other Airbnb categories as much.

[9] See `https://www.stadtentwicklung.berlin.de/planen/basisdaten_stadtentwicklung/lor/` (accessed June 8, 2020).

## 4.3 Descriptives

To better understand our data, we provide various descriptive analyses in this section. We start with some descriptive results on the Airbnb level. We continue at the rental-month level, which is the level of observation in our main analyses. Finally, we present descriptive statistics at the LOR level.

### 4.3.1 Airbnb Descriptives

To illustrate the characteristics of Airbnb in Berlin, we first provide some descriptives on the Airbnb level. Generally, the number of Airbnb listings in the city is increasing. However, the two policy changes in May 2016 and in August 2018 had a clear impact on the number of available Airbnb listings in the city. These impacts are shown in Figure 4.3. This figure shows the number of Airbnb listings available in the city at each of the monthly snapshots, decomposed by the type of Airbnb listings.[10]

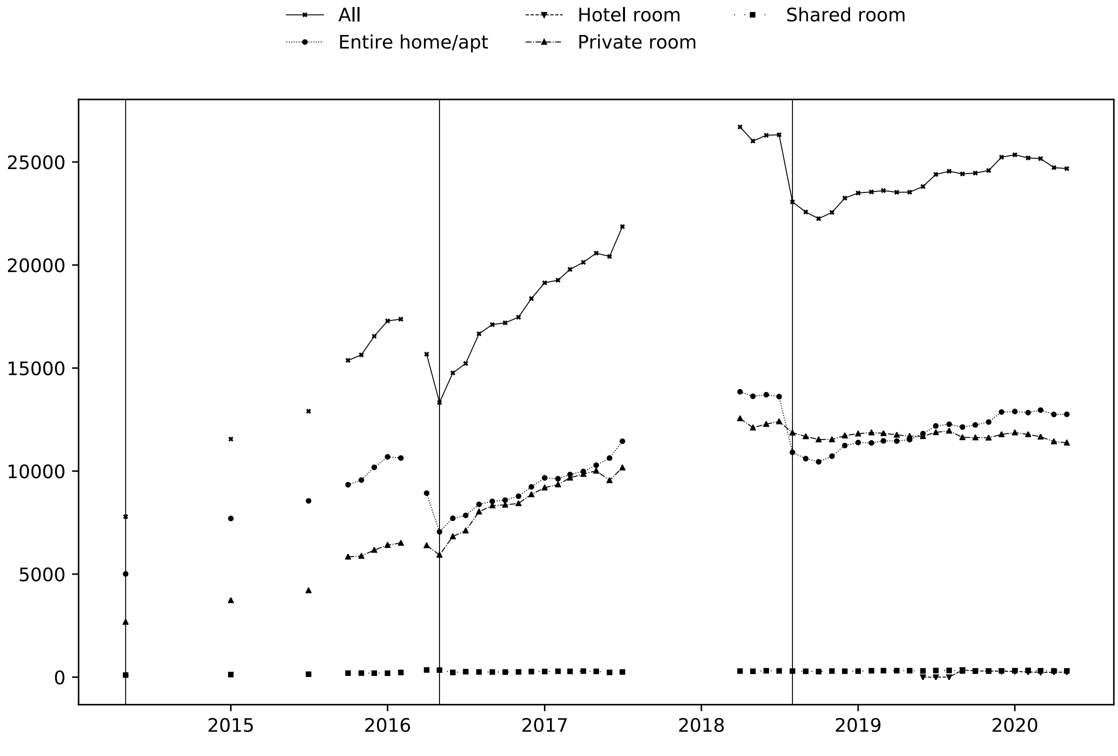



**Figure 4.3:** Number of Airbnb listings in Berlin over time

Over most of the observation period, the number of Airbnb listings in the city is increasing, mostly because the number of entire homes and private rooms in the city is increasing. However, around both reforms, a clear drop in the number of Airbnb listings in the city is visible. Both drops are mostly driven by decreases in the number of entire homes. As a result, while the number of entire homes on Airbnb was approximately double the number of private rooms in the first

[10]An Airbnb listing is part of our data set if it appears as a result when searching for accommodations in Berlin without any date restrictions. This is the case if the listings is set to be available for booking at any time in the future.

monthly snapshot, they are approximately equal toward the end of the observation period. The clear drops around the dates that policy changes were implemented suggest that these changes in the structure of listings on Airbnb are, at least to some extent, driven by policy.

However, the reforms did not just decrease the number of Airbnb listings in the city. The law also affected average availability of listings on the platform. Figure 4.4 shows the average number of days per year that an Airbnb listing was available for booking in each of the monthly snapshots by Airbnb listing type.

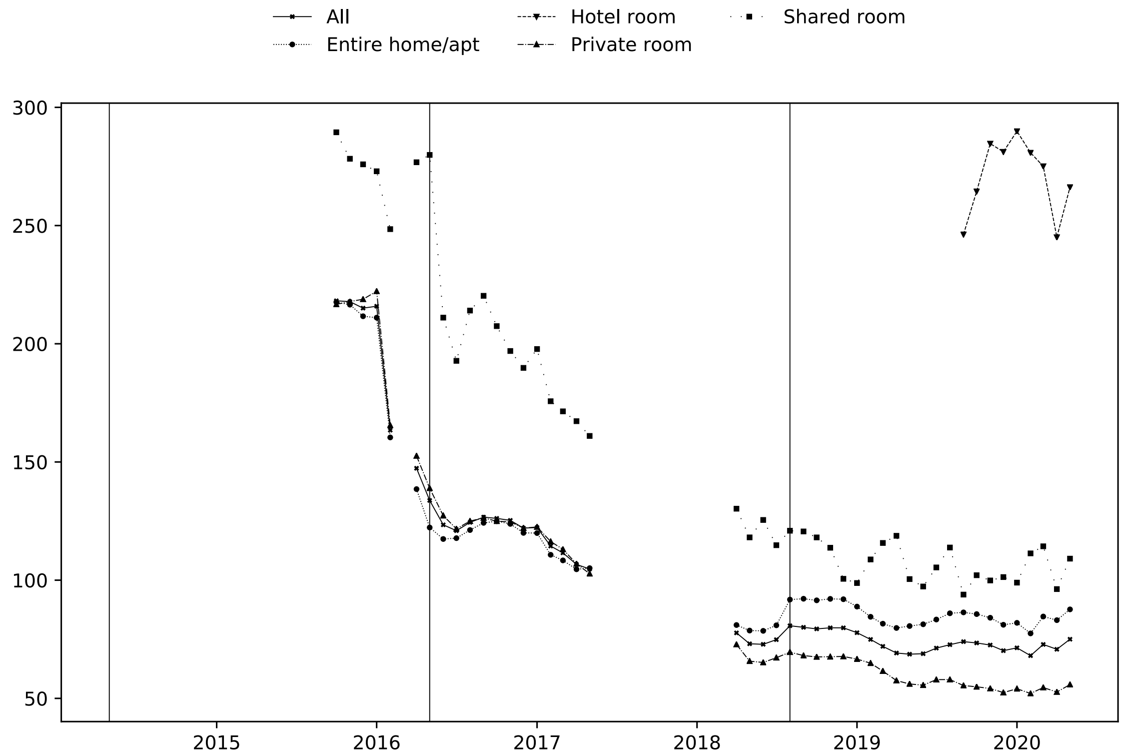



**Figure 4.4:** Mean availability per 365 days by type of Airbnb listing

Leading up to the May 2016 reform, the average availability decreases substantially across all listing types. After May 2016, the mean availability decreases further until the August 2018 reform. Here, the mean availability slightly increases, in particular for entire homes. Interestingly, the new mean in August 2018 seems to lie at an availability of around 90 days per year, which is the cut-off availability below which secondary apartments may also be rented out as short-term rentals following the August 2018 update of the law (see Figure 4.1).

Therefore, while Figure 4.3 suggests that the two policy changes affected Airbnb in Berlin similarly, Figure 4.4 provides more nuance. The May 2016 reform seems to have had a particularly large effect on those Airbnb listings that were available for booking for larger parts of the year. This result is relevant for our main analysis. The main mechanism by which Airbnb is hypothesized to affect rents is that landlords decide to list apartments on Airbnb rather than renting them out long-term (e.g. Yrigoy, 2019). If this mechanism is true, we would expect those Airbnb listings that are listed as short-term rentals for larger parts of the year to impact rents more.

### 4.3.2 Rental Descriptives

In our main analysis, we use disaggregated data at the rental level. Each observation represents one long-term rental apartment in the month in which it was first listed online. Table 4.1 shows selected descriptive statistics calculated for the two samples that we use in our analysis. The "May 2016" sample includes observations from February 2016 to August 2016, while the "August 2018" sample includes data from May 2018 to November 2018. Note that Airbnb data is missing in March 2016.

**Table 4.1:** Rental-level descriptive statistics

| | N | Mean | Std. | Min. | 50% | Max. |
|---|---|---|---|---|---|---|
| *Sample: May 2016* | | | | | | |
| Monthly rent | 23027 | 656.45 | 400.11 | 100 | 542 | 6500 |
| Area (sqm) | 23027 | 70.26 | 30.24 | 11 | 64 | 470 |
| Rent per sqm | 23027 | 9.24 | 2.58 | 3 | 9 | 28 |
| Rooms | 23009 | 2.38 | 0.96 | 1 | 2 | 8 |
| # Airbnb (all, 250m) | 19683 | 18.25 | 29.78 | 0 | 4 | 178 |
| ... available > 180 days | 19683 | 6.62 | 10.55 | 0 | 2 | 75 |
| # Airbnb (entire homes, 250m) | 19683 | 9.97 | 17.04 | 0 | 2 | 119 |
| ... available > 180 days | 19683 | 3.63 | 6.30 | 0 | 1 | 53 |
| # Airbnb (private rooms, 250m) | 19683 | 8.09 | 13.49 | 0 | 2 | 95 |
| ... available > 180 days | 19683 | 2.87 | 4.65 | 0 | 1 | 37 |
| *Sample: August 2018* | | | | | | |
| Monthly rent | 21356 | 755.92 | 466.13 | 122 | 630 | 11000 |
| Area (sqm) | 21356 | 68.88 | 29.74 | 15 | 63 | 551 |
| Rent per sqm | 21356 | 10.86 | 3.30 | 3 | 10 | 43 |
| Rooms | 21332 | 2.34 | 0.94 | 1 | 2 | 9 |
| # Airbnb (all, 250m) | 21356 | 27.66 | 43.87 | 0 | 6 | 299 |
| ... available > 180 days | 21356 | 5.02 | 7.90 | 0 | 1 | 59 |
| # Airbnb (entire homes, 250m) | 21356 | 13.56 | 22.66 | 0 | 3 | 152 |
| ... available > 180 days | 21356 | 2.91 | 5.16 | 0 | 1 | 47 |
| # Airbnb (private rooms, 250m) | 21356 | 13.82 | 22.41 | 0 | 3 | 151 |
| ... available > 180 days | 21356 | 2.04 | 3.28 | 0 | 1 | 24 |

Notes: Descriptive statistics for selected variables on the rental-month level. The upper panel shows the descriptives for the sample surrounding the end of the transition period in May 2016. The lower panel shows the descriptives for the sample surrounding the start of the mandatory registration number display in August 2018.

The average net rent per square meter amounts to 9.24 Euro in the May 2016 sample and 10.86 Euro in the August 2018 sample. On average, each apartment had approximately 18 Airbnb listings within 250 meters in the May 2016 sample. This number increases to about 28 Airbnb listings in the August 2018 sample. In 2016, the largest group of Airbnb listings are entire homes. On average, each rental in the 2016 sample has ten nearby entire homes listed on Airbnb. The entire homes are followed by private rooms with an average of eight listings within 250 meters. In 2018, however, both types of Airbnb listings are approximately equally prominent with apartments having on average 13 of each nearby. Shared rooms make up only a small share of the Airbnb listings. The number of nearby Airbnb listings has large variance and is highly right-skewed. For example, in the May 2016 sample, the median apartment only has two nearby entire homes listed on Airbnb, whereas the apartment with the largest count features 119 nearby entire homes.

Figure 4.5 shows the evolution of the average net rent per square meter over time in the entire city in general and split by Airbnb density in February 2016. We use February 2016 as the base month because it precedes both of our treatment dates as well as the decrease in Airbnb listings before the May 2016 reform (see Figure 4.3). The lines show the average rent for all rentals listed in each month, all rentals at locations that did not have any nearby entire homes on Airbnb in February 2016, all rentals at locations that did have some nearby entire homes in February 2016, and all rentals at locations that had many nearby entire homes in February 2016. To classify the latter two groups, we split those locations with a positive Airbnb count in February 2016 at the median (at ten nearby entire homes). In general, rents are increasing over time. There are some fluctuations and, in particular, the August 2018 reform appears to be followed by a drop in average rents. However, given the variation in these means, it is not clear whether the policy change caused this drop.

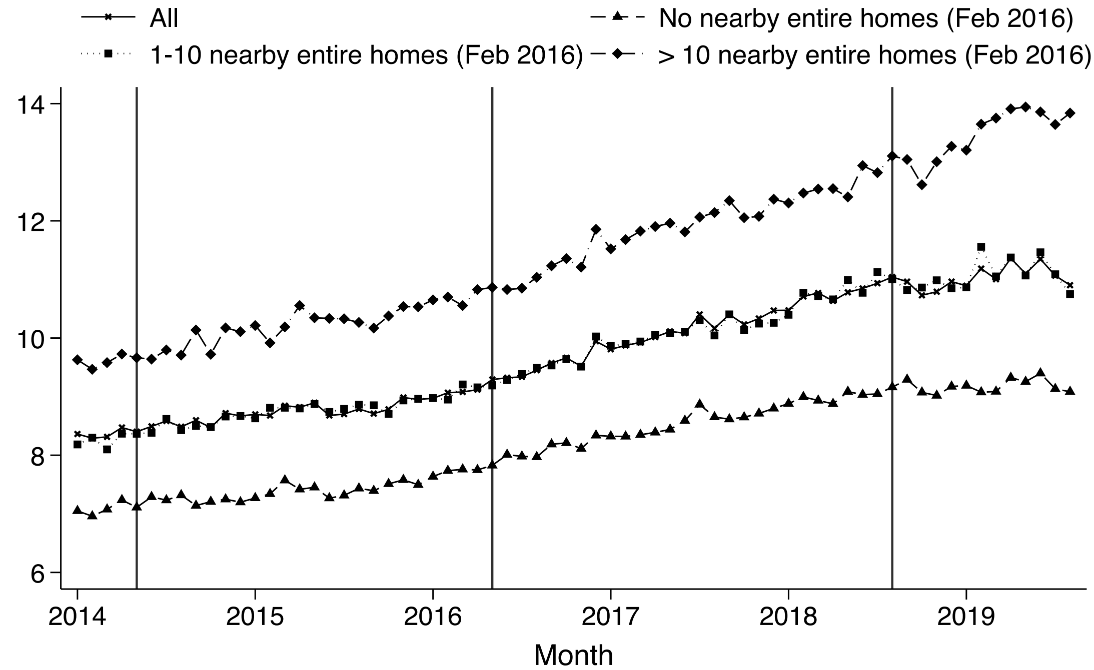



**Figure 4.5:** Average asked rent per square meter of newly listed rentals over time, split by locations with more, fewer, and no entire homes on Airbnb within a 250m distance in February 2016. The vertical lines indicate the policy dates: May 1, 2014; May 1, 2016; and August 1, 2018.

Therefore, in Figure 4.6, we present a more refined view. For this figure, we conduct an analysis inspired by the main specification that Huber *et al.* (2019) employ in an entirely different context. We first need a measure of pre-treatment exposure to Airbnb. For this purpose, we count the number of entire homes within 250 meters, as of February 2016, for the location of each rental in the data. For rental $i$, denote this measure as $abb(Feb2016)_i$. The basic idea then is to interact this measure with a full set of quarter fixed effects and use these interactions in a regression to assess how the discrepancy in rents between areas with fewer and more Airbnb listings in February 2016 changes over time. This regression amounts to estimating the following equation:

$$\begin{aligned} y_{iq} = &\sum_{\tau=Q1,2014}^{Q3,2019} \beta_\tau abb(Feb2016)_i \times \mathbb{1}(q(i)=\tau) + \mathbb{1}(q(i)=\tau) + District_i \times \mathbb{1}(q(i)=\tau) \\ &+ DistrictFE_i + c + \epsilon_{it} \,. \end{aligned} \tag{4.1}$$

Each observation is one rental apartment in the quarter that it was first listed. $y_{iq}$ is the asked rent per square meter. $abb(Feb2016)_i$ denotes the number of entire homes within 250 meters of the location of rental $i$ in February 2016. We interact this cross-sectional measure of pre-treatment Airbnb exposure with a full set of quarter fixed effects. We use the first quarter of 2016 (the quarter containing February 2016) as the base quarter. Further, we include a full set of quarter fixed effects, district fixed effects, as well as the full set of interactions between quarter and district fixed effects. In Figure 4.6, we report the estimates of $\beta_\tau$ for all quarters. These estimates capture the difference in rents between rentals in locations with more and less Airbnb exposure in February 2016, conditional on the fixed effects.

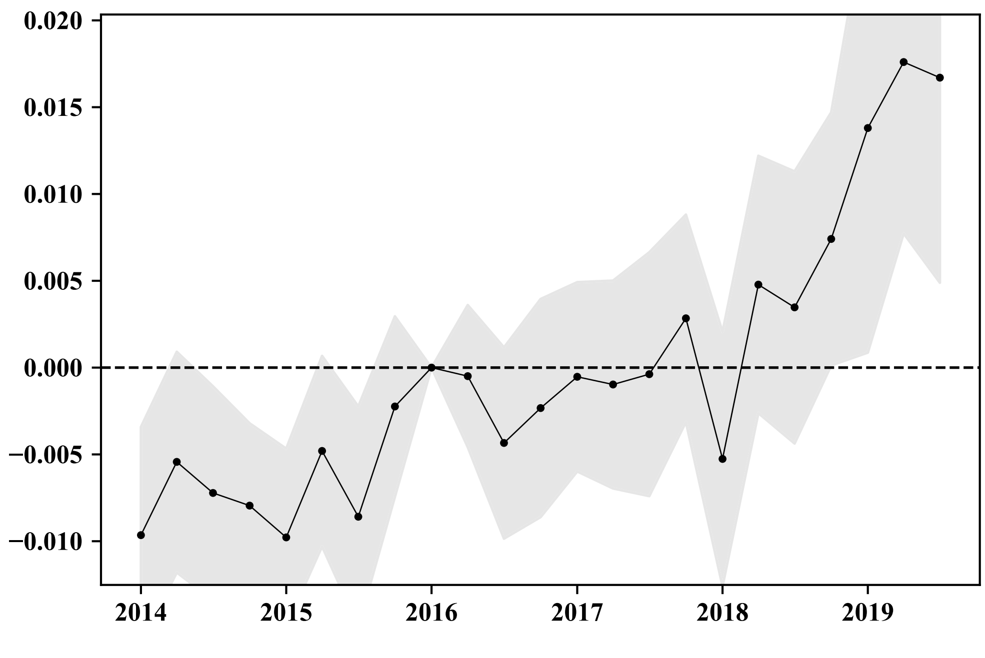



**Figure 4.6:** Point estimates of $\beta_\tau$ from equation (4.1). Shaded areas indicate 95 percent confidence intervals.

The figure indicates that even conditional on this rich set of fixed effects, rents in areas with many Airbnb listings develop differently from those in areas with fewer Airbnb listings. The gap in average rents between low- and high-Airbnb areas seems to be increasing over time. However, the figure also indicates that following the May 2016 reform, there seems to be a temporary break in this trend. This indicates that the law may have had an effect on rents by slowing the growth in areas with more Airbnb listings pre-treatment. However, this effect seems to be transitory as the discrepancy in rents increases again in late 2018.

The plot also suggests that there seem to be other unobserved factors that affect the relationship between Airbnb and rents. These factors make identification, especially in a long-term analysis such as conducted for Figure 4.6, difficult. These insights further motivate why we focus our main analyses on short seven-months windows around the two reforms. With these shorter windows, we can ensure that we better capture the effect of the treatments on Airbnb presence and, consequently, the causal impact of Airbnb on rents.

### 4.3.3 LOR-level Descriptives

In additional analyses, we aggregate the rental-level data set to the LOR level. Each observation represents one LOR in one month. Table 4.2 shows some descriptive statistics on the LOR-level.

**Table 4.2:** LOR-level descriptive statistics

| | N | Mean | Std. | Min. | 50% | Max. |
|---|---|---|---|---|---|---|
| *Sample: May 2016* | | | | | | |
| Rentals per $km^2$ | 2771 | 9.01 | 9.31 | 0 | 6 | 91 |
| Airbnb per $km^2$ | 2375 | 52.65 | 104.37 | 0 | 6 | 786 |
| Entire homes per $km^2$ | 2375 | 28.53 | 58.21 | 0 | 3 | 447 |
| Private rooms per $km^2$ | 2375 | 23.42 | 48.07 | 0 | 3 | 446 |
| Shared rooms per $km^2$ | 2375 | 0.70 | 2.08 | 0 | 0 | 31 |
| *Sample: August 2018* | | | | | | |
| Rentals per $km^2$ | 2738 | 8.42 | 8.60 | 0 | 6 | 77 |
| Airbnb per $km^2$ | 2738 | 82.99 | 158.75 | 0 | 10 | 1068 |
| Entire homes per $km^2$ | 2738 | 40.62 | 80.00 | 0 | 5 | 589 |
| Private rooms per $km^2$ | 2738 | 41.43 | 81.79 | 0 | 5 | 598 |
| Shared rooms per $km^2$ | 2738 | 0.95 | 2.71 | 0 | 0 | 47 |

Notes: Descriptive statistics for selected variables on the LOR-month level. The upper panel shows the descriptives for the sample surrounding the end of the transition period in May 2016. The lower panel shows the descriptives for the sample surrounding the start of the mandatory registration number display in August 2018.

This LOR-level perspective is particularly interesting when analyzing the number of Airbnb listings and rentals in the city rather than rents. On average, there are about eight to nine rentals listed per square kilometer in each month. The average number of Airbnb listings is larger with an average of 53 Airbnb listings per square kilometer around May 2016 and 83 Airbnb listings per square kilometer around August 2018. The numbers of entire homes, private rooms, and shared rooms develop similarly as in the disaggregated data set.

Figure 4.7a shows the average number of rentals per $km^2$ per month in the city by LOR calculated over the entire data set. The rental density appears to be somewhat higher closer to the city center. But there are also various outskirts that have higher rental densities. Figure 4.7b shows a heat map of the average number of entire homes on Airbnb per square kilometer by LOR calculated over the entire data set. Compared to the rentals, entire homes on Airbnb are more concentrated in the city center.


**(a)** Rentals per $km^2$


**(b)** Entire homes on Airbnb per $km^2$

**Figure 4.7:** Heat maps of the average number of long-term rentals and entire homes on Airbnb per square kilometer by LOR for the entire data set

The total number of rentals listed in the city each month fluctuates over time. Figure 4.8 shows the number of newly listed rentals in Berlin over time. Looking at the line for the time around the policy changes (indicated by the second and third vertical lines) may suggest an increase in rental supply following the policy changes. However, as was the case for the average rents, the large fluctuation of the means does not allow for conclusive statements based on this graphical analysis alone.

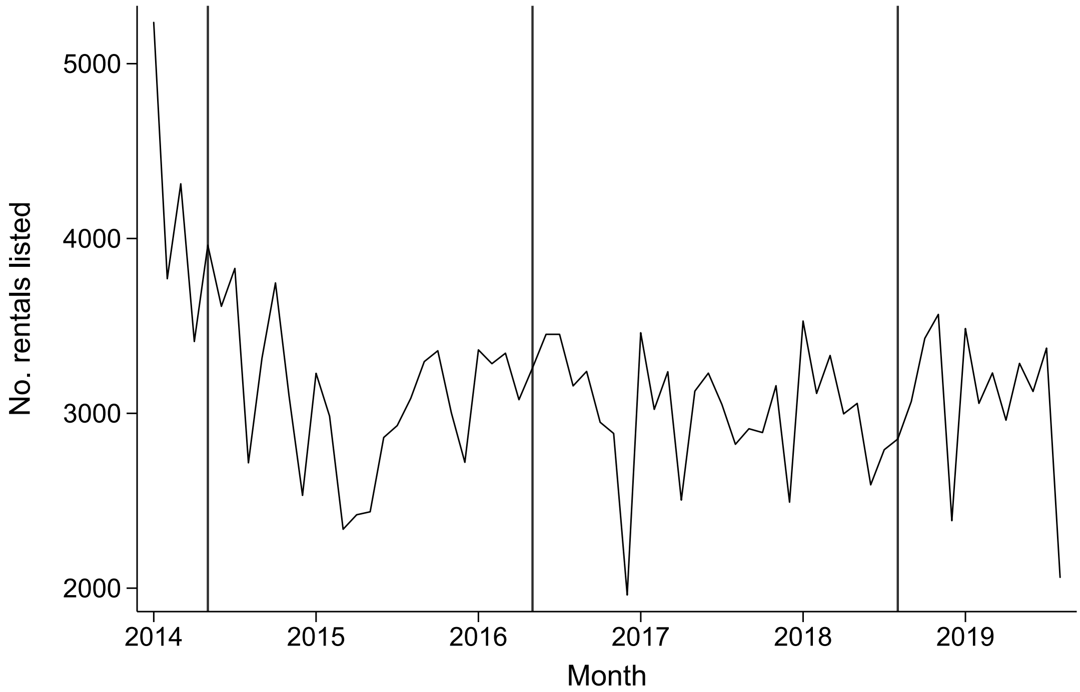



**Figure 4.8:** Newly listed rentals over time. The vertical lines indicate the policy dates: May 1, 2014, May 1, 2016, and August 1, 2018.

We next conduct an analysis of the partial correlation of available Airbnb listings with rents and the supply of long-term rentals in Berlin. Since this analysis is not focused on the impacts of the policy changes, we use the entire sample for which both Airbnb and rent data are available. As Figures 4.3 and 4.5 show, this sample includes data from May 2014 to August 2019. Table 4.3 reports the results of the regressions.

The results show a positive correlation of Airbnb density with rents for all types of Airbnb listings. It also suggests a negative correlation between the density of Airbnb listings with rental density for all types of Airbnb listings. However, the point estimate for shared rooms is quite noisy and not statistically significant at conventional significance levels. These correlations are in line with the intuition for the mechanism of why Airbnb might have an effect on rents: apartments listed as short-term rentals are removed from the long-term rental market, thereby reducing supply, which results in increasing prices.

**Table 4.3:** Airbnb, rents, and rental supply

| | (1) Rent/$m^2$ | (2) Rentals/$km^2$ | (3) Rent/$m^2$ | (4) Rentals/$km^2$ | (5) Rent/$m^2$ | (6) Rentals/$km^2$ |
|---|---|---|---|---|---|---|
| Entire homes per $km^2$ | 0.005*** | -0.031*** | | | | |
| | [0.004; 0.007] | [-0.045; -0.018] | | | | |
| Private rooms per $km^2$ | | | 0.006*** | -0.023*** | | |
| | | | [0.004; 0.008] | [-0.034; -0.012] | | |
| Shared rooms per $km^2$ | | | | | 0.032* | -0.09 |
| | | | | | [0.007; 0.056] | [-0.191; 0.012] |
| N | 15,977 | 15,979 | 15,977 | 15,979 | 15,977 | 15,979 |
| $R^2$ | 0.778 | 0.673 | 0.779 | 0.673 | 0.777 | 0.671 |
| Rent/$m^2$ | 9.820 | 9.820 | 9.820 | 9.820 | 9.820 | 9.820 |

Notes: LOR-level regressions using OLS for all months for which Airbnb data is available. The regressions include LOR fixed effects and a linear time trend. Regressions with rent/$m^2$ as dependent variable also include LOR-month-level mean apartment characteristics. The square brackets show 95 percent confidence intervals. *, **, *** indicate five, one, 0.1 percent significance.

## 4.4 Identification Strategy and Estimation Methods

To analyze the impact of Airbnb on rents, a naive approach is to regress rents on some measure of Airbnb exposure and other covariates:

$$y_{it} = \alpha abb_{it} + x'_{it}\beta + \epsilon_{it} \,. \tag{4.2}$$

In our case, $y_{it}$ is the asked monthly rent per square meter for rental $i$ listed in month $t$, $abb_{it}$ is the number of entire homes listed on Airbnb within 250 meters of the rental, and $x_{it}$ is a vector of exogenous control variables.

### 4.4.1 Identification Strategy

Estimation of Equation (4.2) potentially suffers from endogeneity problems. Arguably most relevant to this particular question are omitted variable bias and reverse causality. Omitted variable bias is an issue if there are unobserved characteristics that affect the attractiveness of an area for both residents as well as tourists and, therefore, would drive both changes in rents as well as in the incentives to host a listing on Airbnb. Reverse causality might be an issue if increasing rents cause residents to list their apartments on Airbnb to subsidize part of their rent. Thus, to establish causality, an identification strategy is needed to address these concerns.

To address the identification concerns, we use the reforms induced by the ZwVbG as quasi-experiments. More specifically, we use these policy changes as instruments in an instrumental variable regression. We also include a potentially large set of covariates to account for apartment characteristics, neighborhood characteristics, as well as postal code level cross-sectional differences in attractiveness. In our main analysis, we therefore estimate

$$y_{it} = \alpha abb_{it} + x'_{it}\beta + \epsilon_{it} \tag{4.3}$$

$$abb_{it} = \gamma \mathbb{1}(t \geq law) + x'_{it}\delta + u_{it} \,, \tag{4.4}$$

where $y_{it}$ is the asked monthly rent per square meter for rental $i$ listed in month $t$, $abb_{it}$ is a measure of Airbnb listings within 250 meters of the rental, $x_{it}$ is a set of exogenous control variables and

$\mathbb{1}(t \geq law)$ is an indicator variable that is one for observations after the date $law$ (we refer to this variable as the "post-dummy"). We run separate regressions using both the end of the transition period on May 1, 2016 as well as the start of the mandatory registration number display on August 1, 2018, as cut-off dates. Thus, the variable $law$ is either May 1, 2016, or August 1, 2018. In the base specification, $abb_{it}$ is the number of nearby entire homes listed on Airbnb but we also present results for different measures to illustrate some nuances of the effect of Airbnb on rents. We also include results in which we interact $\mathbb{1}(t \geq law)$ with postal code fixed effects to allow for more heterogeneity in the first stage.

In order for the policy dummy to be a valid instrument, we need to assume that it is relevant and exogenous conditional on $x_{it}$. We cannot directly test the exogeneity assumption. However, we are confident that there was no parallel policy change that might have affected rents.[11] Figure 4.3 shows a clear drop in the number of Airbnb listings, in particular of entire homes, toward both policy dates. These drops already suggest that the instrument is relevant but we can formally test this using the first-stage regression of Equation (4.4).

In our main analyses, we focus on seven-month time windows around the two reforms. This short-term analysis increases the relevance of our instruments, because as Figure 4.3 suggests, there seem to be long-term trends and other factors that influence the number of Airbnb listings in the city over time. Additionally, the short-term focus helps with identification: Reverse causality is likely less relevant to drive joint changes in rents and Airbnb supply in the short run. Thus, we are mainly concerned with omitted variable bias from unobserved factors that cause both rents and Airbnb supply to change differently in certain regions compared to others. Thus, the direction of the bias of a naive approach, as in Equation (4.2), depends on the direction of both the conditional correlation between changes in Airbnb supply and the unobserved factor as well as the impact of the unobserved factor on changes in rents. Typically, in our context, we would expect unobserved factors to cause a steeper increase in rents in more popular areas, coupled with a larger increase in the supply of Airbnb listings. This would imply a positive bias and the naive approach would overestimate the effect of Airbnb on rents. Note that in our main analysis, we include postal code fixed effects as well as various cross-sectional neighborhood characteristics. Thus, level differences in rents across different regions in Berlin are accounted for. Identification of the Airbnb effect then mostly comes from variation in rents and Airbnb numbers within given regions of the city during our sample period. Since the variation in Airbnb listings in our samples results mostly from a decrease in Airbnb listings, the direction of the bias described above would reverse in our analysis: The decrease in Airbnb listings is likely more pronounced in more popular areas (since those areas have a higher Airbnb density to begin with). Simultaneously, rents in these areas likely still increase more strongly than in less popular areas even if the Airbnb regulation dampens the growth somewhat. Consequently, the naive approach would underestimate the effect of Airbnb on rents. This is exactly what we find in our main analysis.

### 4.4.2 Model Selection

Equation (4.2) illustrates another relevant issue in our context. Because we have rich data on apartment and neighborhood characteristics, we have a potentially large vector $x_{it}$. Including all our covariates would likely result in overfitting of the data, if not dimensionality problems. However, excluding covariates that correlate with both rent and the number of nearby Airbnb

[11] Germany did introduce a rent ceiling for new rentals, but this policy took effect earlier in Berlin, in June 2015. For more details on this rent control policy, please see Mense *et al.* (2017).

listings would introduce an omitted variable bias. Therefore, we need to deal with model selection carefully. To do so, we employ the "double-Lasso" estimators proposed by Belloni *et al.* (2014) for exogenous regressors and Chernozhukov *et al.* (2015) for instrumental variables estimation.

The basic idea of these methods is to use Lasso regression to select those covariates that are most important to explain the dependent variable as well as the explanatory variables of interest. In the most straightforward approach, the union of the sets of selected variables could then be used as covariates in the main regression. Chernozhukov *et al.* (2015) extend this approach. The authors propose an algorithmic estimator that is consistent and robust to small selection errors under certain assumptions. Arguably the most import assumption is that the true underlying model is "approximately sparse." This assumption requires that the true model can be approximated with a small number of variables with only little approximation error. For more details on how we apply the estimator in our context, please refer to Appendix 4.9.1.

This estimator allows us to avoid the usual issues associated with model selection. In principle, as we are agnostic about which of the many covariates to include in the estimation, we let the algorithm decide which specification to use and report.

# 4.5 Results

Before presenting the results of the instrumental variable regression, we start with several preliminary analyses. These analyses are useful for shedding light on the impacts of the reforms on Airbnb and the rental market. We begin with analyses using data aggregated to the LOR-month level. This geographic aggregation allows us to not only investigate prices but also the number of Airbnb listings per $km^2$.

## 4.5.1 Direct Impacts of ZwVbG

The first set of results assesses the direct impacts of the policy changes due to ZwVbG on various outcomes in the Airbnb and rental markets. For this purpose, we focus on the aforementioned seven-month windows around the policy changes. Table 4.4 shows the results of a regression of Airbnb density on the post-dummies for all Airbnb listing types for both policy changes. The estimation for entire homes is equivalent to the first-stage regression we will conduct for the main analysis, except that here, it is conducted at the LOR level. These results confirm formally what an inspection of Figure 4.3 already suggested: Both policy changes resulted in large drops in the number of Airbnb listings in Berlin. These drops are particularly driven by decreases in the number of entire homes listed on Airbnb.

**Table 4.4:** Impact of ZwVbG on Airbnb in Berlin

| | (1) Entire Homes 2016 | (2) Entire Homes 2018 | (3) Private Rooms 2016 | (4) Private Rooms 2018 | (5) Shared Rooms 2016 | (6) Shared Rooms 2018 |
|---|---|---|---|---|---|---|
| Post-dummy | -7.733*** | -9.733*** | -3.839*** | -1.669*** | -0.05 | -0.032 |
| | [-9.656; -5.811] | [-12.180; -7.284] | [-4.803; -2.874] | [-2.432; -0.907] | [-0.138; 0.037] | [-0.121; 0.058] |
| N | 2,375 | 2,738 | 2,375 | 2,738 | 2,375 | 2,738 |
| $R^2$ | 0.971 | 0.977 | 0.976 | 0.997 | 0.894 | 0.958 |
| Rent/$m^2$ | 9.050 | 10.57 | 9.050 | 10.57 | 9.050 | 10.57 |

Notes: LOR-level regressions of the number of Airbnb listings per $km^2$ in each LOR on the post-dummy. The square brackets show 95 percent confidence intervals. *, **, *** indicate five, one, 0.1 percent significance.

Following the May 2016 reform, the number of entire homes per square kilometer decreased by almost eight on average. Table 4.2 shows that the sample mean of entire homes per square kilometer amounts to approximately 29 listings in the May 2016 sample. Thus, this decrease is quite substantial, at around 28 percent. The decrease in entire home density after the August 2018 update is even larger with a reduction of approximately ten listings (ca. 25 percent of the 2018 sample average). Both policy changes also negatively affected the number of private rooms listed on Airbnb. However, these decreases are substantially smaller than for the number of entire homes; even though Table 4.2 shows that the base density of both Airbnb types are quite similar in both samples. The point estimates also suggest a small decrease in shared rooms. However, the estimates are not statistically significantly different from zero at conventional significance levels.

We next investigate the impact of the reforms on the average availability of Airbnb listings. For this purpose, we run the analyses at the Airbnb-month level. Each observation now represents one Airbnb listing in a given month. Table 4.5 shows regressions of the availability per 365 days on the post-dummy for all entire homes in the respective samples. The columns denoted with "Balanced" only include those Airbnb listings that are observed across the entire sample period (i.e. pre- and post-treatment).

**Table 4.5:** Impact of ZwVbG on availability of entire homes on Airbnb in Berlin

| | (1)<br>2016<br>All | (2)<br>2016<br>Balanced | (3)<br>2018<br>All | (4)<br>2018<br>Balanced |
|---|---|---|---|---|
| Post-dummy | -11.53*** | -4.9*** | 7.74*** | -1.037 |
| | [-14.290; -8.781] | [-6.696; -3.104] | [5.710; 9.769] | [-2.108; 0.033] |
| N | 50,381 | 26,621 | 83,456 | 52,351 |
| $R^2$ | 0.090 | 0.130 | 0.105 | 0.150 |

Notes: Airbnb-level regressions of availability per 365 days on the post law dummy for entire homes only. "Balanced" includes only listings that are observed in the entire sample. The square brackets show 95 percent confidence intervals. *, **, *** indicate five, one, 0.1 percent significance.

The results confirm the results from a graphical analysis of Figure 4.4. The May 2016 reform results in a large drop in the mean availability of entire homes on Airbnb. However, the August 2018 update of the law does not. In fact, mean availability even increases for entire homes around the implementation of the update. However, Table 4.5 also provides some new insights. Consider the results in the columns denoted as "Balanced." These results are quite different to those using all entire homes in the sample: The reduction in availability following the May 2016 reform is lower and amounts to only five days on average. The increase following the introduction of the mandatory registration number display in August 2018 vanishes. Combined with the decrease in the number of entire homes due to both policy changes shown in Table 4.4, these results suggest that the decrease in availability in May 2016 and the increase in availability in August 2018 are largely driven by the listings that exit the platform due to the policy changes. In particular, the results suggest that those listings that left the platform in May 2016 were, on average, more active than those remaining on the platform. However, those listings that exited Airbnb toward August 2018 were less active than those remaining on the platform. Considering the different nature of the two policy changes, these different impacts make sense: The end of the transition period in May 2016 effectively marked a ban on professional short-term renting on Airbnb. Therefore, those listings that should be most affected are those that were available for larger parts of the year and, therefore, more likely to be rented out professionally. However, in August 2018, professional short-term renting was already banned but a mandatory registration number display was introduced. This can be seen as the introduction of an additional cost for hosts who were occasionally renting out their apartments in their own absence. While technically allowed to do so, they would have to apply for a permit now to obtain a registration number. For many of these hosts, this additional cost might have been prohibitive and they preferred to delist their apartments from the platform instead.

### 4.5.2 Main Results

We next turn to our main results. Table 4.6 reports the estimates for three specifications for each of the samples. Columns (1) and (4) represent a naive regression of the rent per square meter on the number of nearby entire homes on Airbnb and other covariates. For model selection, the estimator proposed by Belloni *et al.* (2014) is used (denoted as "PDS OLS"). Columns (2) and (5) report the main estimates for an instrumental variable regression of rent per square meter on the number of nearby entire homes on Airbnb using the post-dummies as instruments. Columns (3) and (6) report similar instrumental variable regressions. Here, however, we interact the post-dummy with postal code fixed effects in the first stage to allow for more heterogeneity. Columns (2), (3), (5), and (6) are all estimated using the Chernozhukov *et al.* (2015) estimator (denoted as "Lasso IV").

**Table 4.6:** Main results

| | (1)<br>2016<br>PDS OLS | (2)<br>2016<br>Lasso IV<br>Post | (3)<br>2016<br>Lasso IV<br>Post X Postal | (4)<br>2018<br>PDS OLS | (5)<br>2018<br>Lasso IV<br>Post | (6)<br>2018<br>Lasso IV<br>Post X Postal |
|---|---|---|---|---|---|---|
| *Second Stage* | | | | | | |
| Entire homes (250m) | 0.022*** | 0.101*** | 0.069*** | 0.016*** | 0.04*** | 0.04*** |
| | [0.014; 0.030] | [0.055; 0.146] | [0.035; 0.102] | [0.007; 0.026] | [0.019; 0.061] | [0.019; 0.062] |
| *First Stage* | | | | | | |
| Post-dummy | | -2.788*** | -2.906*** | | -3.26*** | -3.318*** |
| | | [-3.654; -1.923] | [-3.823; -1.988] | | [-4.327; -2.194] | [-4.390; -2.247] |
| N | 19,657 | 19,657 | 19,657 | 21,319 | 21,319 | 21,319 |
| Rent/$m^2$ | 9.260 | 9.260 | 9.260 | 10.86 | 10.86 | 10.86 |
| Selected Xs | 107 | 116 | 108 | 83 | 95 | 93 |

Notes: Rental-month level analyses. Regressions potentially include apartment characteristics, neighborhood characteristics, a quadratic time trend, and postal code fixed effects. The estimation is done using the Belloni *et al.* (2014) estimator for columns (1) and (4) and using the Chernozhukov *et al.* (2015) estimator in the remaining columns. The square brackets show 95 percent confidence intervals. *, **, *** indicate five, one, 0.1 percent significance.

The lower panel of Table 4.6 shows the first-stage results of the IV specifications. For columns (3) and (6), we only show the estimate for the base postal code to economize on space. The results confirm that even on this disaggregated level, the impact of the reforms on the number of Airbnb listings can be seen and is significant both economically as well as statistically. Both reforms reduced the number of entire homes on Airbnb within 250 meters by approximately three units. At averages of ten and 14 units in the two samples (see Table 4.1), this decrease is sizeable also in relative terms.

The upper panel of Table 4.6 reports the main estimates of the corresponding second-stage regressions (or of the main regression in the case of the naive specifications shown in columns (1) and (4)). The results show that additional nearby entire homes on Airbnb increase rents on average. In general, accounting for potential endogeneity using instrumental variables results in larger point estimates, in line with our discussion in Section 4.4.1. Column (2) shows that an additional nearby entire home on Airbnb increases the average asked rent per square meter by approximately ten cents. The estimates using the August 2018 sample are generally lower, suggesting a smaller link between Airbnb and rents. However, the results shown in Table 4.5 suggest that the August 2018 policy change might mostly have affected listings on Airbnb that are unlikely to be relisted as rentals. The results in Table 4.6 do not take into account the availability of nearby Airbnb listings.[12]

Table 4.7 reports the same specifications as Table 4.6 but with one difference: Rather than counting all nearby entire homes on Airbnb, we only count those nearby entire homes that are available for more than 180 out of 365 days. Table 4.1 shows that these listings make up approximately 36 percent of entire homes in the May 2016 sample and only 21 percent in the August 2018 sample.

The lower panel of Table 4.7 shows that the May 2016 reform indeed decreased the number of high availability apartments on Airbnb substantially. It furthers confirms that the August 2018 reform did not have a clear effect on high availability Airbnb apartments. For the second-stage results reported in the upper panel of Table 4.7, these first-stage results suggest that the August 2018 reform estimates need to be interpreted with care as the instruments seem to be weak.

The second-stage estimates suggest that the effect of an additional high availability entire home nearby is larger than the effect ignoring Airbnb availability. Column (2) now suggests a marginal effect of 13 cents per square meter per Airbnb listing. Interestingly, the point estimates using the August 2018 sample are now closer to those using the May 2016 sample, albeit noisier. This result is in line with the first-stage regressions showing that the instrument is weak for high availability apartments in August. That the point estimates are closer to those using the May 2016 sample indicates, however, that counting only the highly active nearby Airbnb listings might be a more relevant measure for the question at hand.

This intuition could also explain the differences in coefficients between the two treatment dates in Table 4.6. In these results, the average decrease in nearby entire homes due to the policy change together with the decrease in average rents in the sample effectively identify the estimate. In the August 2018 sample, however, a large decrease in entire homes did not result in a corresponding decrease in average asked rents. That is why the point estimates are lower than in the May 2016 sample. The results in Table 4.7 now actually suggest that this difference is because in Table 4.6 we used a less relevant measure for nearby Airbnb density. If we instead only count those nearby entire homes that are available for less than 180 days a year, the results suggest smaller effects than those reported in Table 4.6. We report these results in Appendix 4.9.2.

[12] If we include all available covariates and conduct a regular OLS/IV analysis, the point estimates for the May 2016 sample are still positive but smaller and significant only at lower confidence levels. For the August 2018 sample, using all covariates results in statistically insignificant and negative point estimates for the IV methods.

**Table 4.7:** Main results using high availability Airbnb listings only

| | (1)<br>2016<br>PDS OLS | (2)<br>2016<br>Lasso IV<br>Post | (3)<br>2016<br>Lasso IV<br>Post X Postal | (4)<br>2018<br>PDS OLS | (5)<br>2018<br>Lasso IV<br>Post | (6)<br>2018<br>Lasso IV<br>Post X Postal |
|---|---|---|---|---|---|---|
| *Second Stage* | | | | | | |
| Entire homes, | 0.04*** | 0.132*** | 0.095 | 0.05*** | 0.118* | 0.116* |
| available > 180 days (250m) | [0.021; 0.058] | [0.061; 0.202] | [-0.009; 0.199] | [0.023; 0.077] | [0.024; 0.211] | [0.023; 0.209] |
| *First Stage* | | | | | | |
| Post-dummy | | -1.529*** | -1.531*** | | 0.048 | -0.224** |
| | | [-1.963; -1.095] | [-1.967; -1.095] | | [-0.166; 0.262] | [-0.378; -0.070] |
| N | 19,657 | 19,657 | 19,657 | 21,319 | 21,319 | 21,319 |
| Rent/$m^2$ | 9.260 | 9.260 | 9.260 | 10.86 | 10.86 | 10.86 |
| Selected Xs | 77 | 76 | 72 | 91 | 100 | 101 |

Notes: Rental-month level analyses. The Airbnb counts include only those nearby entire homes that are available for more than 180 out of 365 days. All other estimation details are equivalent to those reported in Table 4.6. The square brackets show 95 percent confidence intervals. *, **, *** indicate five, one, 0.1 percent significance.

## 4.6 Robustness

In this section, we present some additional results to address concerns about our identification and check the robustness of our results. First, we address the concern that seasonality might be driving our results. Second, we show how our results react to the use of different Airbnb circle sizes. Third, we discuss how our results change if we focus the analysis on the city center.

### 4.6.1 Seasonality

Because we focus on seven-month windows around the policy changes in our main analyses, we cannot include month fixed effects to take into account seasonality. Therefore, we propose two approaches to discuss the role seasonality plays in our results. First, we run placebo specifications for the first-stage regression to understand if the decrease in Airbnb listings that we assign to the treatments might be due to seasonality. Second, we conduct an analysis in which we first deseasonalize both the rent per square meter as well as the count of nearby entire homes on Airbnb by regressing both on a constant and month fixed effects. We then use the residuals of these regressions and implement the main analyses as reported in Table 4.6.

For the placebo specifications, which can also be viewed as empirical assessments of our identification strategy, we artificially shift the treatment month and use a seven-month window around the artificial treatment month for the analysis. For each of these placebo treatment samples, we run the same analysis as in columns (2) and (5) of Table 4.6 and investigate the first-stage results. Figure 4.9 reports the coefficient estimates for the post-dummy for each of these artificial treatment samples. The y-axis shows the point estimates and the x-axis shows the corresponding placebo treatment months. Because of the gaps in the Airbnb data, we cannot conduct the analysis for all months for which we have rent data. More specifically, we only include those treatment months for which we have at least two months of Airbnb data prior to and two months after the treatment.

The results show that the drop in Airbnb numbers after May 2016 and August 2018 clearly stem from the policy changes, as none of the other coefficients are equivalent in absolute size. Toward the end of the transition period in May 2016, there seems to be some anticipation, with the drop actually already starting in February 2016. If we use February 2016 as the treatment month, the estimated effects in the second stage are similar (see Appendix 4.9.3).

These results also indicate that it is difficult to interpret the second-stage results from such a

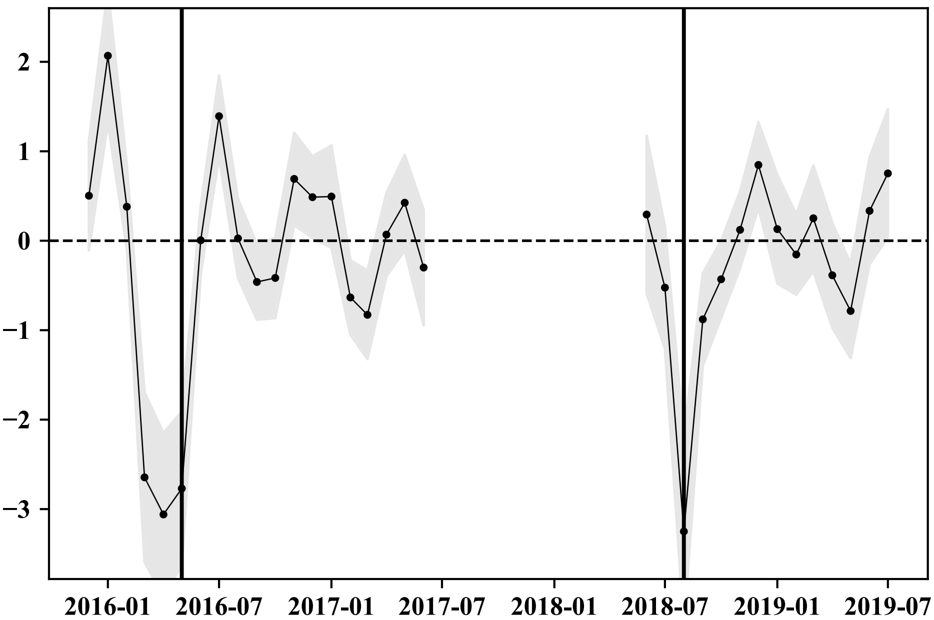



**Figure 4.9:** First-stage regressions of a Lasso IV of rent per $m^2$ on entire homes on Airbnb (250m) using a shifting post-dummy as an instrument. The samples always consist of the six months surrounding the treatment month and the treatment month. The plot shows the coefficient estimates on the y-axis and the posited treatment month on the x-axis. The shaded areas show 95 percent confidence intervals.

placebo exercise. Figure 4.9 shows that the post-dummy is a weak instrument for many of the placebo treatment months away from the actual policy months. Consequently, it is not clear that the corresponding second-stage estimates would be unbiased. Therefore, we propose a different method to assess the impact of seasonality on our second-stage results instead. We deseasonalize rent per square meter as well as the number of nearby entire homes on Airbnb by regressing both on month fixed effects and a constant. We then use the residuals from these regressions in our main specifications. To account for the additional variation from the deseasonalization preceding the main analysis, we use the bootstrap for inference. For more details on our bootstrapping procedure, please refer to Appendix 4.9.4.

Table 4.8 reports the results of this exercise. The results show that seasonality does not seem to drive our results in the May 2016 sample. The results using IV for the August 2018 sample are no longer statistically significantly different from zero. This could again be a result of the fact that the August 2018 reform did not affect highly active Airbnb listings as much as the May 2016 reform.

## 4.6.2 Airbnb Circle Size

The choice to use a distance of 250 meters to count the number of Airbnb listings nearby a rental is ad-hoc. Choosing the size of the circle to draw around each rental is subject to a trade-off. On the one hand, we expect Airbnb listings that are closer to a rental to have a larger impact on

**Table 4.8:** Main results using deseasonalized variables

| | (1)<br>2016<br>PDS OLS | (2)<br>2016<br>Lasso IV<br>Post | (3)<br>2016<br>Lasso IV<br>Post X Postal | (4)<br>2018<br>PDS OLS | (5)<br>2018<br>Lasso IV<br>Post | (6)<br>2018<br>Lasso IV<br>Post X Postal |
|---|---|---|---|---|---|---|
| Entire homes (250m) | 0.018***<br>[0.011; 0.024] | 0.108***<br>[0.065; 0.153] | 0.046***<br>[0.023; 0.067] | 0.012***<br>[0.007; 0.018] | 0.007<br>[-0.021; 0.034] | -0.002<br>[-0.016; 0.014] |
| Draws | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 |

Notes: Rental-month level analyses. We deseasonalize the rent per square meter and the number of nearby entire homes on Airbnb by regressing them on month fixed effects and a constant first. We then use the resulting residuals in the estimations. All other estimation details are equivalent to those reported in Table 4.6. For inference, we draw bootstrap samples before conducting the deseasonalization and estimation. The square brackets show 95 percent confidence intervals calculated as the 2.5 and 97.5 sample percentiles of the bootstrapped coefficient estimates. *, **, *** indicate five, one, 0.1 percent significance, calculated using the percentile bootstrap. For more details on our bootstrapping procedure, please refer to Appendix 4.9.4.

rents on average. Choosing a larger circle size would therefore dilute the estimated effects. On the other hand, the location of Airbnb listings is anonymized by the platform and randomly shifted by up to 150 meters. Choosing a smaller circle size makes the measurement error resulting from this anonymization more severe. Our main results use circles with a radius of 250 meters. In the following, we report the same estimations, but using circle sizes with radii of 500 and 1000 meters.

**Table 4.9:** Main results using different circle sizes

| | (1)<br>2016<br>PDS OLS | (2)<br>2016<br>Lasso IV<br>Post | (3)<br>2016<br>Lasso IV<br>Post X Postal | (4)<br>2018<br>PDS OLS | (5)<br>2018<br>Lasso IV<br>Post | (6)<br>2018<br>Lasso IV<br>Post X Postal |
|---|---|---|---|---|---|---|
| Entire homes (500m) | 0.004*<br>[0.000; 0.007] | 0.038***<br>[0.024; 0.053] | 0.025***<br>[0.013; 0.036] | 0.004*<br>[0.000; 0.008] | 0.007<br>[-0.001; 0.015] | 0.006<br>[-0.001; 0.013] |
| Selected Xs | 115 | 107 | 106 | 61 | 70 | 61 |
| Entire homes (1000m) | 0.0<br>[-0.001; 0.002] | 0.008***<br>[0.004; 0.012] | 0.007***<br>[0.003; 0.010] | 0.001<br>[-0.001; 0.002] | 0.002<br>[-0.000; 0.004] | 0.002<br>[-0.000; 0.004] |
| Selected Xs | 103 | 96 | 102 | 41 | 94 | 94 |
| N | 19,657 | 19,657 | 19,657 | 21,319 | 21,319 | 21,319 |
| Rent/$m^2$ | 9.260 | 9.260 | 9.260 | 10.86 | 10.86 | 10.86 |

Notes: Rental-month level analyses. We use the number of nearby entire homes on Airbnb within 500 meters and 1000 meters as our Airbnb measure. All other estimation details are equivalent to those reported in Table 4.6. The square brackets show 95 percent confidence intervals. *, **, *** indicate five, one, 0.1 percent significance.

Table 4.9 shows that larger circle sizes result in smaller effect sizes. This is in line with our intuition that closer Airbnb listings should have a larger impact on rents. At the same time, the confidence intervals become tighter with larger circle sizes, which is in line with the idea that larger circle sizes allow more precise Airbnb counts as the measurement error introduced due to the anonymization of the exact location becomes less severe. In all specifications, however, the effect of Airbnb on rents is positive, although the exact point estimates do vary.

### 4.6.3 City Center versus the Entire City

Figure 4.7b shows that most Airbnb listings are located within the city center. This fact might lead to concerns that very little variation in the number of Airbnb listings might be driving our results if we include districts in which the number of Airbnb listings does not change (much) over time. The upper panel of Table 4.10 therefore shows the results of conducting the main analysis for only those listings within the city center. We define the city center as all areas that are within the so-called “S-Bahnring,” a circular train track that surrounds the inner city. The point estimates are indeed smaller but still clearly positive and significant for the May 2016 sample. The estimates for the August 2018 sample are no longer significant. This pattern might again be a result of high-availability Airbnb listings having a larger impact on rents.

**Table 4.10:** Main results for the city center

| | (1)<br>2016<br>PDS OLS | (2)<br>2016<br>Lasso IV<br>Post | (3)<br>2016<br>Lasso IV<br>Post X Postal | (4)<br>2018<br>PDS OLS | (5)<br>2018<br>Lasso IV<br>Post | (6)<br>2018<br>Lasso IV<br>Post X Postal |
|---|---|---|---|---|---|---|
| Entire homes on Airbnb (250m) | 0.013**<br>[0.005; 0.021] | 0.046**<br>[0.011; 0.081] | 0.032*<br>[0.001; 0.064] | 0.006<br>[-0.003; 0.015] | 0.02<br>[-0.012; 0.052] | 0.015<br>[-0.017; 0.048] |
| Selected Xs | 27 | 33 | 29 | 35 | 37 | 36 |
| Entire homes, available > 180 days (250m) | 0.023**<br>[0.006; 0.040] | 0.068*<br>[0.016; 0.120] | 0.061*<br>[0.002; 0.121] | 0.022<br>[-0.002; 0.046] | 0.022<br>[-0.131; 0.176] | 0.042<br>[-0.099; 0.184] |
| Selected Xs | 26 | 31 | 28 | 25 | 36 | 33 |
| Rent/$m^2$ | 10.78 | 10.78 | 10.78 | 12.88 | 12.88 | 12.88 |
| Selected Xs | 26 | 31 | 28 | 25 | 36 | 33 |

Notes: Rental-month level analyses. We use only those listings located in the city center (defined as the area within the S-Bahnring). All other estimation details are equivalent to those reported in Table 4.6. The square brackets show 95 percent confidence intervals. *, **, *** indicate five, one, 0.1 percent significance.

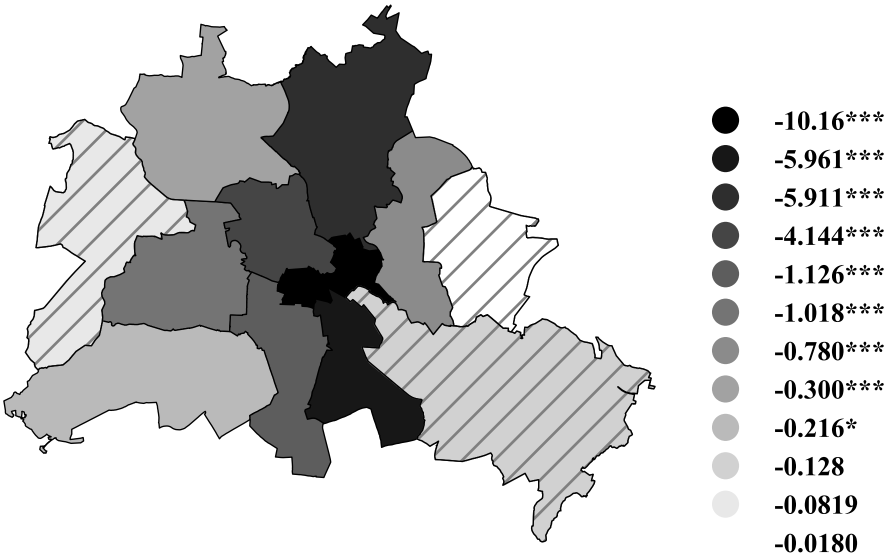
If we focus on only those entire homes that are available for more than 180 days per year, the point estimates for the May sample are again larger. We report the results in the lower panel of Table 4.10. However, the point estimates in the August 2018 are still not statistically significantly different from zero, although they are positive. This non-significance most likely is a result of the August 2018 reform not causing sufficient variation in the availability of Airbnb listings. Combined with the reduced sample size due to restricting the analysis on the city center, we can no longer identify an effect in the August 2018 sample.

## 4.7 Heterogeneity

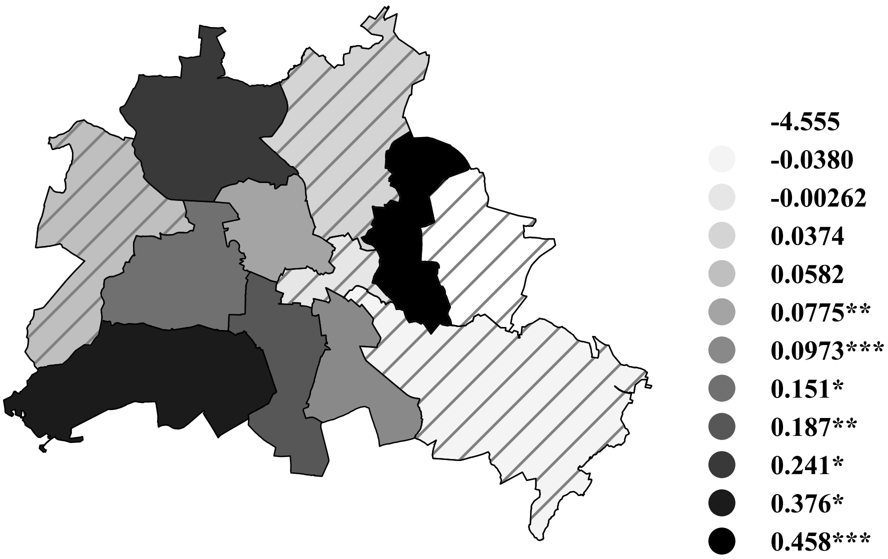
Thus far, we have only considered average effects. However, the effect of Airbnb on rents is most likely heterogeneous across different regions and apartments. Therefore, we next investigate effect heterogeneity across different regions of the city as well as different types of rentals.

### 4.7.1 Geographic Heterogeneity

We first turn to a look at the heterogeneous effects of Airbnb on rents across the city. To do so, we stratify the sample by district and conduct the same analysis as reported in column (2) in Table 4.6 using the May 2016 sample for each of the sub-samples. To first understand the heterogeneity of the effect of the May 2016 reform, Figure 4.10a shows the first-stage estimates of the post-dummy coefficient for each district. Unsurprisingly, the central districts that had a higher density of Airbnb listings pre-reform (see Figure 4.7b) also experience the largest decreases in entire homes due to the law. For every district, the point estimates are negative. However, note that not all of the point estimates are actually statistically significantly different from zero. Districts for which the point estimates are not statistically significantly different from zero at the 95 percent confidence level are overlaid with diagonal stripes. Additionally, *, **, *** indicate five, one, 0.1 percent significance in the legend.




**(a)** First-stage estimates for the post-dummy




**(b)** Second-stage estimates of the Airbnb effect

**Figure 4.10:** District-stratified Lasso IV regressions of rent per square meter on the number of entire homes on Airbnb (250m) using the post-dummy as an instrument for the May 2016 sample. The specification is identical to the one reported in column (2) of Table 4.6. *, **, *** indicate five, one, 0.1 percent significance. Districts for which the coefficient is not statistically significant at the 95 percent confidence level are marked with diagonal stripes.

Figure 4.10b shows the second-stage estimates of the effect of Airbnb on rents corresponding to

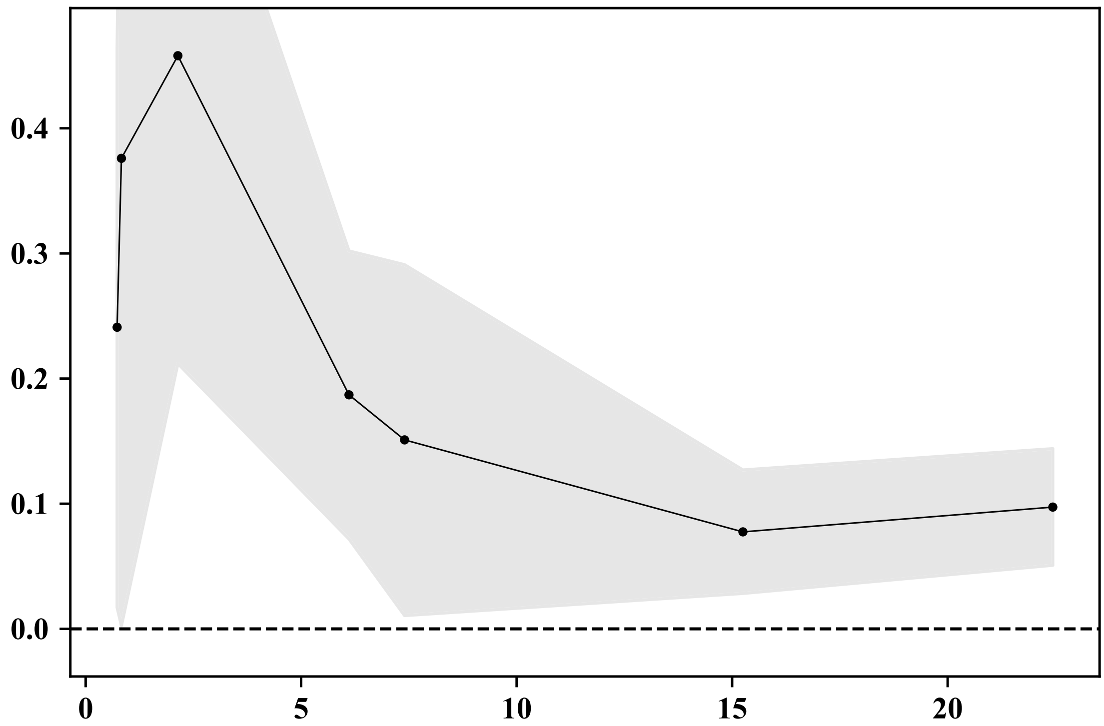
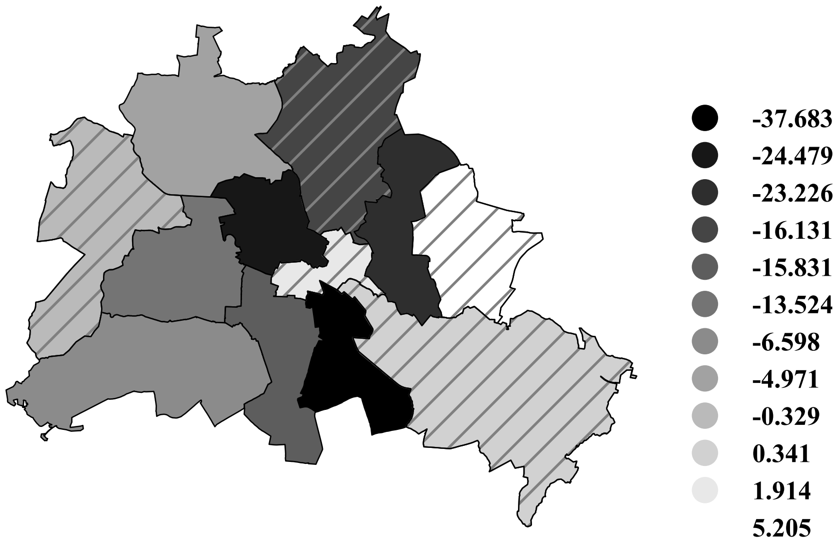



**Figure 4.11:** Estimates by Airbnb distribution: Statistically significant estimates (95 percent confidence level) from Figure 4.10b plotted against the mean number of entire homes on Airbnb (250m) by district. Shaded area shows 95 percent confidence intervals.

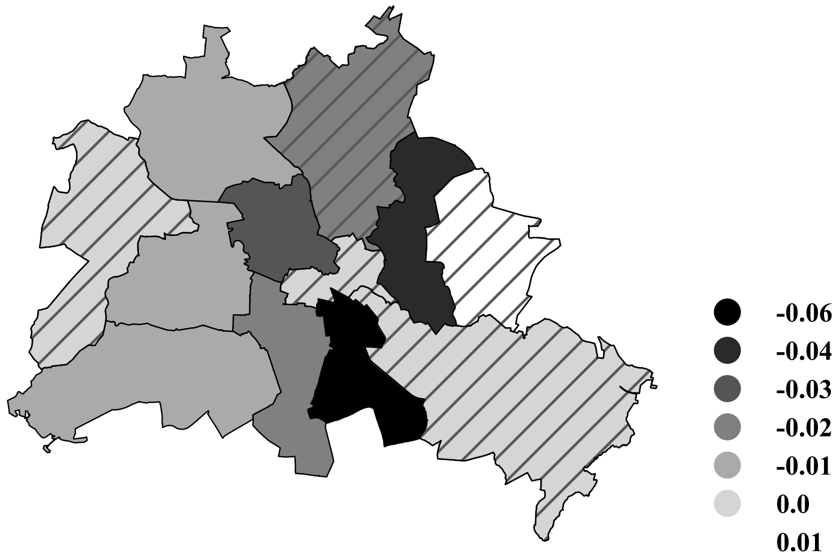
the first-stage results presented in Figure 4.10a. Note that this figure suggests negative effects for three districts. However, the estimates are not statistically significant at conventional confidence levels for any of these districts. All of the statistically significant estimates suggest a positive effect of Airbnb on rents. Further, the estimates suggest effect heterogeneity across districts with increases in rent per square meter ranging from eight to 46 cents per additional nearby entire home.

Against our intuition, these estimates suggest that the marginal effect of additional Airbnb listings is larger in the outskirts of the city. In fact, plotting the estimates that are statistically significant at the 95 percent confidence level against the district-level mean number of entire homes listed on Airbnb in a 250 meter circle shows that districts with a lower Airbnb density tend to be subject to larger marginal Airbnb effects. This result is shown in Figure 4.11. Although the line is not monotonically decreasing, it does suggest decreasing marginal effects of Airbnb on rents. One possible explanation for this pattern might be that with high Airbnb density, the negative externalities on residents (e.g. noise at night) might become large enough to reduce residential demand in the area. Exploring these channels would be interesting but is outside the scope of this paper. These results are not conclusive of course. First, the estimates are noisy and the insight is based on the results for seven districts only. Second, the difference in mean Airbnb density is based on different districts here. There might be some district-level characteristics that drive both Airbnb density and the marginal effect of Airbnb on rents. Nevertheless, these results are novel to the literature and give future research a direction to explore further in more suitable settings.

Note that Figure 4.10a suggests that more central districts experienced larger decreases in the




**(a)** Implied cumulative decrease in monthly rents




**(b)** Implied cumulative decrease in monthly rents relative to average rent

**Figure 4.12:** Cumulative decrease in monthly rents. Products of the estimates from Figure 4.10a, Figure 4.10b, as well as the average apartment size (Figure 4.12a) and relative to average monthly rent by district (Figure 4.12b). Districts for which the second-stage coefficient is not statistically significant at the 95 percent confidence level are marked with diagonal stripes (these districts include all districts for which the first-stage coefficient is not statistically significant at the 95 percent confidence level).

number of entire homes listed on Airbnb as a result of the May 2016 reform. At the same time, however, Figure 4.10b suggests that these same districts are subject to lower marginal effects of

Airbnb on rents. When considering how the May 2016 reform may have impacted rents across different districts, these two effects imply opposite relationships between more central and outer districts: On the one hand, the decrease in the number of Airbnb listings was larger in the city center, on the other hand, the marginal effect of each of these Airbnb listings on rents is lower. To give an idea about what our estimates imply for the cumulative effect of the May 2016 reform on rents, we multiply the estimates shown in Figures 4.10a and 4.10b. Further, we calculate the average size of rentals for each district and multiply it with the product of the two stages. The result can be interpreted as the monthly rent that an average apartment in each district saved due to the May 2016 reform. The results of this exercise are shown in Figure 4.12a. The results suggest that the effect of the law on rents is heterogeneous across districts and amounts to up to approximately 38 Euros per month in the more popular Airbnb districts. Figure 4.12b shows the results of Figure 4.12a as a percentage of the average rent calculated by district. The cumulative effect of the law is calculated to decrease total monthly rent up to six percent.

### 4.7.2 Heterogeneity by Apartment Type

We next turn to an investigation of the heterogeneity of the Airbnb effect by different types of apartments. In particular, we are interested in whether rentals with a different number of rooms are affected differently by nearby Airbnb listings. For this purpose, we split the sample of rentals by the number of bedrooms. Note that most apartments have integer number of bedrooms, but sometimes apartments also have, for example, half rooms. In these cases, we round down. For example, we group all one-bedroom apartments together with all one-and-a-half-bedroom apartments. Further, we combine all apartments with four or more bedrooms in one group. For these sub-samples, we then conduct the same analysis as in Table 4.6. Table 4.11 reports the results of this exercise.

The results suggest some effect heterogeneity. In particular, column (3) suggests somewhat larger effects for apartments with more bedrooms. In general, however, the results do not suggest a clear-cut pattern of heterogeneity across apartments with different numbers of bedrooms.

## 4.8 Conclusion

The impact of short-term rental platforms such as Airbnb on the long-term rental market is a politically relevant and much discussed topic. Even though cities around the world have already introduced policies to regulate short-term rental platforms, empirical evidence studying the causal link between the spread of Airbnb and outcomes in the housing market only started emerging in 2017. Establishing and quantifying the causal impact of Airbnb on rents is challenging.

We contribute to this literature by providing the first causal study on this topic for Berlin, Germany. We further contribute to the small group of papers that use a policy change as a natural experiment to address the inherent endogeneity in a causal analysis of Airbnb and rents. We use legislation that was passed by the City of Berlin (zwvbg) to curb the mis-use of apartments as full-time short-term rentals. For the analysis, we combine data on asked rents, Airbnb in Berlin, and neighborhood characteristics from various data sources. To account for observable confounders, we use a rich set of potential regressors that explain both the number of nearby Airbnb listings as well as the rents of specific apartments. We select the relevant regressors using the so-called double-Lasso estimators proposed by Belloni *et al.* (2014) and Chernozhukov *et al.* (2015) in order to remain agnostic in the model selection process.

**Table 4.11:** Main results stratified by number of bedrooms

| | (1)<br>2016<br>PDS OLS | (2)<br>2016<br>Lasso IV<br>Post | (3)<br>2016<br>Lasso IV<br>Post X Postal | (4)<br>2018<br>PDS OLS | (5)<br>2018<br>Lasso IV<br>Post | (6)<br>2018<br>Lasso IV<br>Post X Postal |
|---|---|---|---|---|---|---|
| $0 \leq Bedrooms < 2$ | | | | | | |
| Entire homes (250m) | 0.025*** | 0.073 | 0.069 | 0.018* | 0.012 | 0.011 |
| | [0.012; 0.038] | [-0.002; 0.149] | [-0.006; 0.145] | [0.002; 0.033] | [-0.073; 0.098] | [-0.074; 0.096] |
| N | 3,422 | 3,422 | 3,422 | 4,010 | 4,010 | 4,010 |
| Rent/$m^2$ | 9.800 | 9.800 | 9.800 | 11.44 | 11.44 | 11.44 |
| Selected Xs | 30 | 32 | 32 | 37 | 28 | 26 |
| $2 \leq Bedrooms < 3$ | | | | | | |
| Entire homes (250m) | 0.014** | 0.056*** | 0.061*** | 0.02*** | 0.042** | 0.04** |
| | [0.005; 0.023] | [0.024; 0.088] | [0.034; 0.089] | [0.009; 0.030] | [0.016; 0.068] | [0.015; 0.065] |
| N | 8,855 | 8,855 | 8,855 | 9,622 | 9,622 | 9,622 |
| Rent/$m^2$ | 9.120 | 9.120 | 9.120 | 10.70 | 10.70 | 10.70 |
| Selected Xs | 116 | 116 | 110 | 53 | 76 | 65 |
| $3 \leq Bedrooms < 4$ | | | | | | |
| Entire homes (250m) | 0.026** | 0.096* | 0.072** | 0.025*** | 0.031 | 0.036* |
| | [0.010; 0.042] | [0.023; 0.168] | [0.021; 0.123] | [0.014; 0.035] | [-0.044; 0.106] | [0.001; 0.071] |
| N | 5,258 | 5,258 | 5,258 | 5,545 | 5,545 | 5,545 |
| Rent/$m^2$ | 9.010 | 9.010 | 9.010 | 10.55 | 10.55 | 10.55 |
| Selected Xs | 59 | 49 | 56 | 41 | 41 | 44 |
| $Bedrooms \geq 4$ | | | | | | |
| Entire homes (250m) | 0.03** | 0.067* | 0.112** | 0.016 | 0.066* | 0.085** |
| | [0.009; 0.050] | [0.006; 0.128] | [0.029; 0.195] | [-0.001; 0.033] | [0.004; 0.127] | [0.022; 0.147] |
| N | 2,122 | 2,122 | 2,122 | 2,142 | 2,142 | 2,142 |
| Rent/$m^2$ | 9.590 | 9.590 | 9.590 | 11.32 | 11.32 | 11.32 |
| Selected Xs | 44 | 61 | 54 | 17 | 45 | 40 |

Notes: Analysis as in Table 4.6, but with rentals stratified by the number of bedrooms. The square brackets show 95 percent confidence intervals. *, **, *** indicate five, one, 0.1 percent significance.

In a first step, we find that the policy changes resulted in marked decreases in the number of Airbnb listings in Berlin. Furthermore, in particular when the law first fully took effect in May 2016, not only did the number of Airbnb listings (particularly entire homes listed on Airbnb) decrease, but the average number of days per year that an Airbnb listing was available for booking also dropped substantially. This change in availability was different when the law was updated in August 2018. While this update also resulted in a strong decrease in the number of entire homes listed on Airbnb in Berlin, it mostly affected less active listings.

Our results suggest that one additional entire home listed on Airbnb within 250 meters of a rental apartment increases the average asked monthly rent per square meter by at least seven cents. Focusing on highly active Airbnb listings, the estimated impact is even larger at 13 cents per additional Airbnb. We find that this effect is heterogeneous across districts. We document suggestive evidence that the effect might be larger in districts with a lower Airbnb density.

Thus, our paper confirms the results of previous research that finds that rents increase with increasing Airbnb popularity in the US (Horn and Merante, 2017; Koster *et al.*, 2018; Barron *et al.*, 2020) and in Europe (Garcia-López *et al.*, 2020). However, our results also document some new results with regard to the heterogeneity of the effect. In particular, we find that the more highly available Airbnb listings are those that seem to be most relevant for the rental market.

For policy-makers, these results suggest that regulation of the short-term market aiming to ease the burden on the long-term rental market should take into account carefully which types of short-term rental listings will be affected. While the initial introduction of the regulation in Berlin strongly reduced the number of high availability listings in Berlin, its update mostly affected less active listings. With regard to rents, the initial introduction (or, rather, the end of the transition period after the law was introduced) was then most likely more effective in relieving the rental market. The August 2018 update, however, affected less active Airbnb listings more and, thusly, might have hurt hosts who profit from occasionally hosting on Airbnb but did not, in fact, withdraw apartments from the long-term rental market. These nuances are new to the literature and should be explored more by future research to further inform the policy debate.

# 4.9 Appendix

## 4.9.1 Our Application of Chernozhukov *et al.* (2015)

We briefly outline the algorithm of the estimator applied to our problem here. For a more detailed discussion, please refer to Chernozhukov *et al.* (2015). Consider the moment condition

$$E[(\tilde{\rho}^{y}_{it} - \tilde{\rho}^{abb}_{it}\alpha)\tilde{\nu}_{it}] = 0\,, \tag{4.5}$$

where $\tilde{\rho}^{y}_{it} = y_{it} - x'_{it}\theta$, $\tilde{\rho}^{abb}_{it} = abb_{it} - x'_{it}\vartheta$, and $\tilde{\nu}_{it} = x'_{it}\delta + \gamma\mathbb{1}(t \geq law) - x'_{it}\vartheta$.

Chernozhukov *et al.* (2015) show that this moment condition is valid around the true parameter values, even for small deviations from the true parameter values. Because of this result, the moment condition is "immune" to small selection errors. This moment condition corresponds to an exogeneity assumption when regressing $\tilde{\rho}^{y}_{it}$ on $\tilde{\rho}^{abb}_{it}$ using $\tilde{\nu}_{it}$ as an instrument. Therefore, the authors propose to estimate exactly this instrumental variable regression in order to obtain an estimate for $\alpha$, the coefficient of interest.

Chernozhukov *et al.* (2015) propose to obtain the sample equivalents of the necessary expressions using the following algorithm:

1. Conduct a first-stage regression of $abb_{it}$ on $\mathbb{1}(t \geq law)$ and $x_{it}$ and denote the corresponding coefficients as $\hat{\gamma}$ and $\hat{\delta}$. Obtain predicted Airbnb counts using $\hat{abb}_{it} = \hat{\gamma}\mathbb{1}(t \geq law) + x'_{it}\hat{\delta}$.

2. Conduct a regression of $y_{it}$ on $x_{it}$ and denote the corresponding coefficient as $\hat{\beta}$.

3. Conduct a regression of $\hat{abb}_{it}$ on $x_{it}$ and denote the corresponding coefficients as $\hat{\upsilon}$.

4. Calculate $\hat{\rho}^{y}_{it} = y_{it} - x'_{it}\hat{\beta}$, $\hat{\rho}^{d}_{it} = \mathbb{1}(t \geq law) - x'_{it}\hat{\upsilon}$, and $\hat{\nu}_{it} := \mathbb{1}(t \geq law)\hat{\gamma} + x'_{it}\hat{\delta} - x'_{it}\hat{\upsilon}$. Use IV regression of $\hat{\rho}^{y}_{it}$ on $\hat{\rho}^{d}_{it}$ using $\hat{\nu}_{it}$ as an instrument to obtain $\hat{\alpha}$.

The authors propose to use either Lasso or Post-Lasso (OLS using variables previously selected by Lasso) to run the three regression steps and obtain the parameter estimates. Asymptotically, the choice of the estimator makes no difference. We use Lasso for the estimation.

Chernozhukov *et al.* (2015) show that standard inference methods for IV regression are valid for $\hat{\alpha}$. As mentioned above, the authors show that using the IV regression of the transformed prediction errors in step 4 amounts to using a moment restriction that makes the estimator robust to small model selection mistakes.[13]

[13] The authors also discuss that if perfect model selection were possible, then the transformation were not necessary. Instead, it would be valid to use the union of the $x_{it}$ that were selected in steps 1 and 2, together with the instrument $\mathbb{1}(t \geq law)$, in a regular IV framework.

### 4.9.2 Main Results for Low Availability Airbnb Listings

**Table 4.12:** Main results using low availability Airbnb listings only

| | (1)<br>2016<br>PDS OLS | (2)<br>2016<br>Lasso IV<br>Post | (3)<br>2016<br>Lasso IV<br>Post X Postal | (4)<br>2018<br>PDS OLS | (5)<br>2018<br>Lasso IV<br>Post | (6)<br>2018<br>Lasso IV<br>Post X Postal |
|---|---|---|---|---|---|---|
| Entire homes, available $<$ 180 days (250m) | 0.022***<br>[0.012; 0.033] | 0.066***<br>[0.029; 0.104] | 0.064***<br>[0.027; 0.102] | 0.018**<br>[0.006; 0.029] | 0.043*<br>[0.003; 0.082] | 0.032**<br>[0.011; 0.054] |
| N | 19,657 | 19,657 | 19,657 | 21,319 | 21,319 | 21,319 |
| Rent/$m^2$ | 9.260 | 9.260 | 9.260 | 10.86 | 10.86 | 10.86 |
| Selected Xs | 127 | 134 | 124 | 78 | 80 | 96 |

Notes: Rental-month level analyses. The Airbnb counts include only those nearby entire homes that are available for less than 180 out of 365 days. All other estimation details are equivalent to those reported in Table 4.6. The square brackets show 95 percent confidence intervals. *, **, *** indicate five, one, 0.1 percent significance.

### 4.9.3 Main Results for February 2016 Treatment

**Table 4.13:** Main results for February 2016 treatment

| | (1)<br>2016<br>PDS OLS | (2)<br>2016<br>Lasso IV<br>Post | (3)<br>2016<br>Lasso IV<br>Post X Postal |
|---|---|---|---|
| Entire homes (250m) | 0.016***<br>[0.009; 0.024] | 0.098**<br>[0.036; 0.160] | 0.112***<br>[0.069; 0.155] |
| N | 18,687 | 18,687 | 18,687 |
| Rent/$m^2$ | 9.070 | 9.070 | 9.070 |
| Selected Xs | 128 | 146 | 131 |

Notes: Rental-month level analyses. Same specifications as reported in Table 4.6, except that we use February 2016 as the treatment month. The square brackets show 95 percent confidence intervals. *, **, *** indicate five, one, 0.1 percent significance

### 4.9.4 Bootstrapping Procedure

For the results reported in Table 4.8, we draw bootstrap samples before conducting the deseasonalization to account for the variation introduced by the deseasonalization procedure. Note that although the main analysis is focused around the short time windows around the May 2016 and August 2018 reforms, we need larger time windows to be able to include month fixed effects in our deseasonalization. Therefore, for the deseasonalization, we use the entire data set of rents and Airbnb listings available to us. This implies that we are drawing bootstrap samples from a larger sample that we end up using in the main estimation in each bootstrap iteration. Let $N$ denote the total number of rentals in our data. In each bootstrap iteration $s$, we then follow the following steps:

1. Draw $N$ rentals with replacement from the full data set.
2. Use this bootstrap sample and regress

$$y_{it} = \alpha + \beta MonthFE_t + \epsilon_{it}\,, \tag{4.6}$$

   where $y_{it}$ is either the rent per square meter or the measure of nearby Airbnb listings of rental $i$ listed in month $t$. This regression yields coefficient estimates $\hat{\alpha}$ and $\hat{\beta}$.
3. Calculate $\hat{\epsilon}_{it} = y_{it} - \hat{\alpha} - \hat{\beta} MonthFE_t$ for both variables.
4. Use only those rentals out of the $N$ bootstrap rentals that are within the sample time window around May 2016 or August 2018. Denote the number of corresponding rentals as $N_s$. Note that while $N$ is constant for all bootstrap iterations, $N_s$ can vary. For these $N_s$ rentals, we run the main specifications replacing rents per square meter and the Airbnb measure with the corresponding estimated residuals.

As point estimates, we report the results from the estimation using the original sample. Denote these point estimates as $\hat{\beta}_0$. For inference, for each coefficient, we save all estimates from each of the bootstrap iterations. To calculate 95 percent confidence intervals, we simply use the 2.5 and 97.5 percentiles of the sample distribution of these estimates. To calculate p-values, we shift all of these estimates by their mean to center them around zero. We then calculate the probability to obtain estimate $\hat{\beta}_0$ given that the true parameter distribution is the distribution of estimates centered around zero. To do so, we calculate the percentage of parameter estimates that are below $0 - |\hat{\beta}_0|$ or above $0 + |\hat{\beta}_0|$.

# Chapter 5

# Ratings and Strategic Pricing: Evidence from Airbnb

**Bibliographic information: This chapter has not been published yet. Maximilian Schaefer is the only author of this chapter.**

## 5.1 Introduction

Review systems are arguably among the most critical components of peer-to-peer platforms. Building trust between anonymous market participants is at the core of the success of the sharing economy. Platforms like Airbnb, eBay, and TaskRabbit crucially rely on their rating mechanisms to generate and sustain interaction between the various market sides. Ideally, reviews should convey an accurate signal about the quality of a product/service in order to avoid negative user-experiences that would otherwise threaten the most valuable asset of a platform: the size of its user-base.

In the economic literature, the impact of ratings on revenues and related performance measures receives considerable attention. Several studies establish a robust positive correlation between a seller's rating and prices (Teubner *et al.*, 2017), revenues (Luca, 2016), and quantities (Livingston, 2005). These results are usually understood as indicative of the value of a seller's reputation. In this regard, it is interesting to notice that this literature views ratings as a proxy for reputation rather than quality.

Therefore, taken at face value, the results of the aforementioned literature imply that ratings likely capture additional features beyond the pure "quality" of a product/ service. An important question in this respect is to understand (a) what other factors (besides quality) determine a seller's reputation, and (b) whether the seller can strategically affect these factors to reap the benefits ensuing from a better reputation.

In this paper, we use transaction and rating data from the Airbnb platform to analyze the impact of prices on ratings. We find robust evidence that higher prices negatively affect ratings, which is consistent with the notion that travelers do not simply rate the quality of an accommodation but rather the quality net of the price. This finding suggests that there is some room for hosts to use pricing as a mean to strategically influence ratings. Consistent with this, we provide empirical evidence suggesting that hosts charging low prices when entering the market can generate above-average initial ratings. Interestingly, these hosts, after charging low-entry prices, are subsequently able to increase prices without suffering a comparable rating depreciation as do hosts who charge higher prices from the beginning.

This empirical contribution complements the theoretical work of Stenzel *et al.* (2020), who analyze sellers' pricing incentives in a theoretical framework where prices are explicitly allowed to influence ratings. Their analysis reveals an interesting trade-off: higher prices decrease the net valuation of a product, which tends to depreciate ratings. At the same time, high prices cause customers with a high valuation of the product to purchase, which tends to increase ratings. The results of the present analysis suggest that the first effect dominates the latter in the setting of this paper. The presented findings are consistent with the work of Luca and Reshef (2020), who find that prices negatively impact star ratings on Yelp.

In our results section, we first analyze the extent to which salient rating changes correlate with the performance of hosts on the Airbnb platform. Using a fixed effect approach, we show that the overall rating of an accommodation at the beginning of a month is positively correlated with the performance of a listing in the following month. According to our estimates, an additional half star rating (which ranges from one to five stars in incremental steps of 0.5 stars) increases monthly revenue by 40 Euros. The positive revenue effect seems to be driven by an increase in the number of bookings rather than by an increase in prices, which appear not to be correlated with the overall rating. The nature of our data also allows us to apply a regression discontinuity design which confirms the findings obtained with the fixed-effect approach.

In a second step, we analyze the correlation between prices and ratings. Using a fixed effects approach, we estimate that a ten percent price increase during a given month reduces average overall rating at the end of that month by approximately 0.15 stars. We also find that prices are positively correlated with the probability of observing negative rating changes. Additionally, prices are negatively correlated with the probability of observing positive rating changes. In terms of magnitude, our regression coefficients suggest that a ten percent price increase increases the probability of observing a negative change in rating by 50 percentage points and reduces the probability of observing a positive change in rating by 80 percentage points. Interestingly, other rating measures, which refer to specific non-price aspects and are routinely collected by Airbnb, appear not to be impacted much by prices.

The only other rating measure that is consistently negatively correlated with prices is the value-for-money rating, which asks travelers to rate the value-for-money aspect of the accommodation. Thus, our results indicate that only the overall rating and the value-for-money rating are negatively correlated with prices. This can be seen as indicative evidence that consumers factor prices into the overall rating. Since the overall rating is the most salient rating displayed on the Airbnb website, the results suggest that prices are crucial for a host's overall reputation.

Finally, we gather suggestive evidence that hosts can build a comparative advantage in reputation by charging low prices when entering the market. Interestingly, host with low initial prices manage to sustain the resulting initial reputation advantage despite increasing prices subsequently. While the increase in prices leads to a depreciation of ratings, which is consistent with our findings, this decrease is not enough to offset the initial advantage in reputation that low entry-prices generate in comparison to high entry-prices.

Our paper relates to the literature on the value of reputation (Eaton *et al.*, 2002; Houser and Wooders, 2006; Livingston, 2005; Luca, 2016; Lucking-Reiley *et al.*, 2007; Melnik and Alm, 2002; Resnick *et al.*, 2006; Teubner *et al.*, 2017). The paper sheds light on the determinants of a seller's reputation (Hubbard, 2002; Jin and Leslie, 2009; Luca and Zervas, 2016; Mayzlin *et al.*, 2014; Proserpio and Zervas, 2017) and specifically highlights the role of prices (Cabral and Li, 2015;

Zegners, 2019). The paper also broadly relates to the literature analyzing biases in consumer ratings (Bolton *et al.*, 2013; Li and Xiao, 2014; Li *et al.*, 2016).

The remainder of the article is organized as follows. In Section 5.2, we describe the various data sources used for the analysis. In Section 5.3, we provide an overview of how individual ratings are aggregated on Airbnb. In Section 5.4, we briefly analyze the relationship between ratings and several performance measures. In Section 5.5, we analyze the impact of prices on ratings. In Section 5.6, we investigate the pricing and ratings dynamics of listings entering the market. Section 5.7 concludes.

# 5.2 The Data

## 5.2.1 Data Description

We combine three different data sources for the analysis. The period of observation spans the entire year of 2017. The geographical area covered is the city of Paris, France. In the following, we briefly describe each of the three different data sources in turn.

**Transaction data:** The transaction level data are provided by Airdna[1]. This data source allows for determining, for each listing and for each date in 2017, if a listing was booked on that particular date and the daily rate charged. The data also contain reservation identifiers that allow for determining for how long an accommodation was booked by a particular individual. Thus, it allows for determining the exact time-span of a single reservation. For a more detailed discussion of the Airdna Data see Schaefer and Tran (2020).

**Rating data:** For the rating information, we use web-scraped data from inside Airbnb.[2] In contrast to the transaction data, the rating data only contain monthly information. At the beginning of each month in 2017, we observe the aggregated ratings for each listing. Listings are rated on seven different categories: (i) the overall rating, (ii) value-for-money, (iii) cleanliness, (iv) check-in comfort, (v) location, (vi) accuracy, and (vii) communication. The rating information we observe is an aggregate of the all the individual ratings a listing received. Thus, each monthly rating observation can be considered as an update of the aggregate ratings.[3]

All rating categories range from one to five stars in incremental steps of 0.5 stars. The most salient rating for the customer is the overall rating, which is prominently displayed on the top of the webpage describing a listing. The overall rating asks the traveler to rate the "overall experience" of a stay. The other categories ask for specific aspect of the stay and are largely self-explanatory from their name.[4] The accuracy rating (vi) asks whether the description on the Airbnb website is compatible with the experience of the traveler. The communication rating (vii) asks whether the host was easily reachable and helpful when the guest needed assistance or help.

A noteworthy aspect that we will exploit for a regression discontinuity approach is the fact that we observe the overall rating category on a more granular level than the customer. While customers only see the star rating, we observe an integer valued measure ranging from 20 to 100. A star rating of one corresponds to [20,24], a star rating of 1.5 corresponds to [25,34], and so on.

[1]See `https://www.airdna.co` (last accessed: September 17, 2020)

[2]see `http://insideairbnb.com/get-the-data.html` (last accessed: September 17, 2020)

[3]The review process is described in more detail in Appendix 5.8.1.

[4]See `https://www.airbnb.com/help/article/1257/how-do-star-ratings-work-for-stays` for Airbnb's description of the various subcategories (last accessed: September 17, 2020)

A five star rating corresponds to [95, 100]. For the other rating categories, we observe the same star rating as the customer, which is encoded by integers ranging from 2 to 10: 2 corresponds to one star, 3 to one and a half stars, and so on. Thus, for the overall rating, we are able to capture non-salient rating changes.

**Text Review Data:** In addition to the star rating, guests are prompted to review a listing by leaving a text describing their experience. The text review data are also provided by Inside Airbnb and contain, for each listing, the timestamp and the exact text of an individual review. Unfortunately, the corresponding information on the star rating of an individual review is not collected. Nevertheless, the timestamps contained in the text review data are useful for our purposes, as will be explained in more detail in subsection 5.2.2.

### 5.2.2 Matching Price and Review Data

Transaction and text review data are available on the daily level, rating data only on the monthly level. Ideally, we would like to observe transaction-level (daily) information on the different rating categories. It should also be emphasized that the rating information refers to aggregate ratings with an unknown aggregation mechanism that we discuss in Section 5.3. In an ideal setting, we would like to observe individual ratings for each stay in an accommodation to assess the impact of prices on individual ratings. In the following, we explain how we match the data in an attempt to come as close as possible to the ideal setting.

The Airdna data allow us to determine when a stay in a particular accommodation ended. The text review data allow us to determine when a listing was rated. From the rating data, we can observe how the aggregate rating changes on a monthly basis. Imagine, for the sake of exposition, that we observe the aggregate ratings for two consecutive months. If we observe one reservation followed by a text review in between the dates for which we observe updates of the aggregate ratings, we can reasonably assume that any changes in the aggregate rating of a listing is attributable to the reviewed reservation.

Naturally, since we match different data sources, we must resort to some approximations. We employ the following method: a reservation is considered as reviewed if we observe a text review occurring within 14 days after the last day of the reservation.[5] For all reservations that end between two aggregate ratings updates $t$ and $t + 1$, we assign the average price of these reservations to the aggregate rating information in $t + 1$. This allows us to analyze how aggregate rating updates change when average monthly reservation prices change. In the scenario where we have only one reviewed reservation between two ratings updates, changes in the aggregate ratings can be tied to the price of one reservation. This is essentially the closest we can get to transaction level rating information.

From 630,000 unique reviews that we observe, we are able to match 570,000 to reservations observed in the Airdna data.[6]

[5] Guests can leave a review within 14 days after their stay, see `https://www.airbnb.com/help/article/995/how-long-do-i-have-to-write-a-review-after-a-trip-has-ended` (last accessed: September 17, 2020).

[6] In total, we observe 1.4 million reservations in the Airdna data. This implies that approximately 30 percent of all guests leave a review.

# 5.3 Aggregation of Individual Ratings

Airbnb does not disclose information on the exact functioning of its rating mechanism. From visiting forums on which Airbnb hosts exchange experiences, it is obvious that Airbnb does not make transparent to the hosts how individual ratings are aggregated.[7] Thus, this section aims to shed light on the likely aggregation mechanism of individual listings. We analyze how frequently aggregate ratings change as a function of the total number of ratings a listing received. Since we do not observe transaction-level star ratings, it is important to gain an understanding of the aggregation mechanism. If, for instance, the aggregate rating is calculated as the simple average of all past individual ratings, a new incoming rating will not affect the aggregate rating for listings that were already rated frequently.

Figure 5.1 shows the relative frequency of changes in the aggregate overall rating as a function of the number of reviews. The left panel displays the results for the non-salient granular overall rating measure, the right panel refers to salient changes in the overall star rating. Figure 5.1 focuses only on observations with less than 50 reviews. We use the convention that for the first review the frequency of changes is equal to 100 percent. From the left panel of figure 5.1, we see that based on all listings for which we observe ten reviews, 60 percent experienced a change in the aggregate non-salient rating compared to the aggregate rating after nine reviews. The relative frequency of changes falls off steadily. For the salient rating measure, we naturally observe fewer changes. The pattern observed for the salient overall quality rating (right panel of Figure 5.1) also holds for the other rating subcategories, the corresponding figures are found in Appendix 5.8.2.

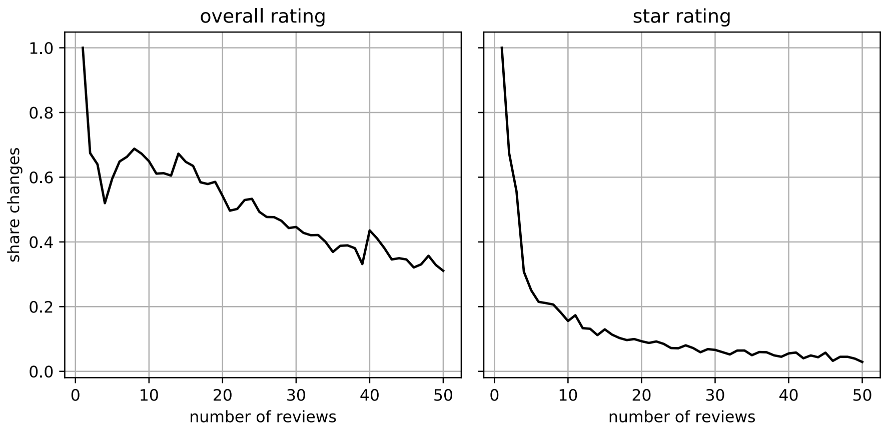



**Figure 5.1:** Relative frequency of overall rating changes as a function of the number of reviews

The declining rate of changes speaks in favor of a simple averaging of the individual ratings. Thusly, the impact of newly incoming individual ratings is more pronounced for listings with only a limited number of existing reviews. In the remainder of the analysis, we control for the number of existing reviews of a listing when assessing the impact of prices on aggregate ratings. Furthermore, we provide robustness checks for results obtained using all observations by restricting the sample to listings with a small number of reviews. For listings with a small number of reviews, the impact

[7]For an interesting forum discussion among Airbnb hosts, see `https://airhostsforum.com/t/how-exactly-is-the-star-rating-calculated/14575` (last accessed: September 18, 2020).

of individual ratings on the aggregate ratings is more pronounced. Therefore, focusing on such listings should allow us to approximate the impact of prices on individual reviews more precisely.

Another important issue concerns the relationship between the various rating categories. For instance, it could be that the overall rating is calculated as an average of the rating subcategories. However, discussion among Airbnb hosts indicates that the overall rating is *not* aggregated from the subcategories.[8] In principle, it is possible that a reviewer gives five stars for all rating subcategories but only four stars for the overall rating. This is also consistent with the way ratings are collected, since the overall rating is the first that a guest submits (see Appendix 5.8.2). There seems to be no mechanical "by-design" relationship between the subcategories and the overall rating.

## 5.4 The Impact of Ratings on Performance

Before analyzing the impact of prices on ratings, we first briefly assess whether our data confirm the general finding in the literature that positive star ratings improve the revenues of sellers. To do so, we run the following regression specification:

$$y_{it} = \beta_0 + \beta_1 rating_{it-1} + X_{it} + \mu_i + \lambda_t + \epsilon_{it} \tag{5.1}$$

In the following, $y_{it}$ is the average price, the number of unique bookings, the total days booked, or the revenue observed for a listing $i$ in a particular month $t$. $rating_{it-1}$ is the overall aggregate rating at the beginning of the month $t$. The subscript $t-1$ emphasizes that we focus on the last available aggregate rating observed before measuring the performance measures in $t$.

$X_{it}$ are time-varying control variables. In our baseline specification, $X_{it}$ consists only of the number of days a listing was available for booking in a particular month. The more days a listing is available for booking, the more likely it is to be booked. Hence, it is important to control for the availability when analyzing the impact of the overall rating. In the Appendix, we present several robust checks in which we add controls to the matrix $X$. For the overall rating, we use the salient star measure.

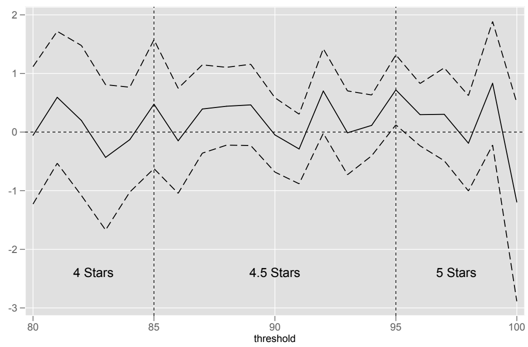
**Table 5.1:** Impact of overall star rating on several performance measures

| | (1)<br>price | (2)<br>bookings | (3)<br>days booked | (4)<br>revenue |
|---|---|---|---|---|
| rating: overall | 0.271 | 0.115*** | 0.428*** | 41.25*** |
| | (0.145) | (0.0184) | (0.0428) | (5.794) |
| days available | 0.0140 | 0.150*** | 0.813*** | 71.11*** |
| | (0.00847) | (0.000997) | (0.00235) | (0.498) |
| Constant | 87.02*** | -0.676*** | -8.736*** | -848.8*** |
| | (1.354) | (0.174) | (0.407) | (56.80) |
| Observations | 170290 | 170290 | 170290 | 170290 |

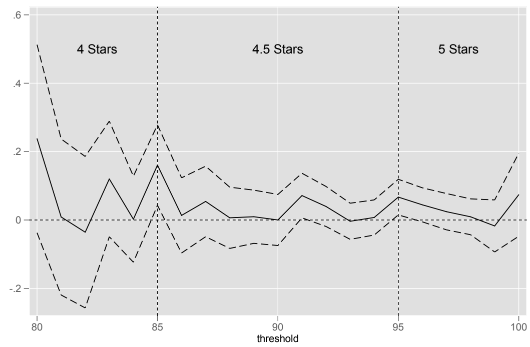
Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5.1 shows the results obtained from estimating Equation 5.1 for the price, the number

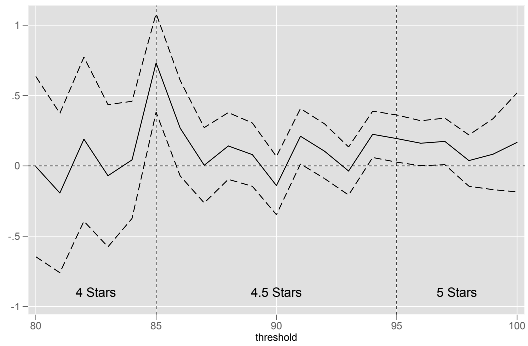
[8]See, for example, `https://community.withairbnb.com/t5/Hosting/5-stars-in-all-categories-but-4-star-stay/td-p/693470` (last accessed: September 18, 2020).



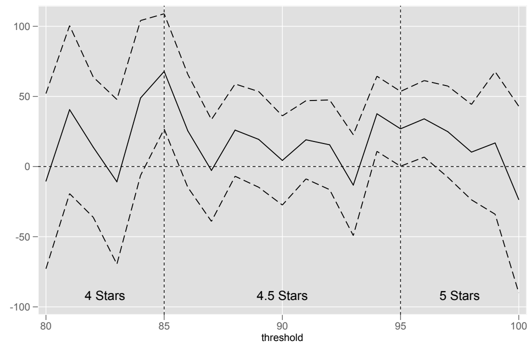

**(a)** $\beta_1$ for $y_{it}$ = price




**(b)** $\beta_1$ for $y_{it}$ = number of bookings




**(c)** $\beta_1$ for $y_{it}$ = number of days booked




**(d)** $\beta_1$ for $y_{it}$ = revenue

**Figure 5.2:** Results of the regression discontinuity design for different thresholds and dependent variables. Dashed lines represent the 95 % confidence intervals

of bookings, the total number of days booked, and the revenue. All regressions control for month and listing fixed effects, with standard errors clustered on the listing level. We find a statistically significant and positive coefficient for the number of bookings, the total number of days booked, and the revenue. For prices, we find no significant estimate. According to the above estimates, an increase in the overall rating by half a star, increases monthly revenue by 41.25 Euros.[9] The results suggest that the increase in revenue is mostly driven by an increase in demand rather than price adjustments.[10]

The fact that we observe the overall rating on a more disaggregated level than the consumers allows us to apply a regression discontinuity design to analyze the impact of ratings on performance. The data allow us to compare listings which are almost identical with respect to the overall non-salient rating but which differ with respect to the salient rating. For example, listings with an overall rating of 94 and 95 are almost identical with respect to the overall non-salient rating

[9] The star rating is re-scaled such that an increase in rating by one unit of the re-scaled measure corresponds to an increase in the rating by half a star. Note that the salient rating measure changes in steps of half a star.

[10] In Appendix 5.8.3, we present several robustness checks. First, we add several control variables such as the price of the listing (when the price is not the dependent variable), the number of reviews, and measures capturing the demand state in the neighborhood of an accommodation. Additionally, we also present regression results when including the other rating subcategories. The results presented in Table 5.1 are robust to the inclusion of various control variables. In particular, the price never seems to be affected by ratings.

measure but differ in the salient measure. To analyze the impact of salient rating thresholds on performance, we run the following regression:

$$y_{it} = \beta_0 + \beta_1 \mathbb{1}_{bw}(rat_{it} > \tau^\star) + X_{it} + \mu_i + \lambda_t + \epsilon_{it} \tag{5.2}$$

In Equation 5.2, $bw$ denotes the bandwidth chosen for the regression discontinuity design and $\tau^\star$ denotes the threshold. $\mathbb{1}_{bw}(rat_{it} > \tau^\star)$ is an indicator variable that takes the value one if the non-salient rating exceeds the threshold. For $bw = 0.5$ and $\tau^\star = 80.5$, $\beta_1$ captures the difference in $y_{it}$ between listings with a rating of 81 and 80.

In the following, we fix $bw = 0.5$ and let $\tau^\star = \{80.5, 81.5, 82.5, ...., 99.5\}$. Thus, we estimate the coefficient $\beta_1$ for listings which are adjacent in the non-salient overall rating. For the threshold values of $\tau^\star = \{84.5, 94.5\}$ the difference in the non-salient measure will result in changes in the salient rating measure, otherwise not.

Figure 5.2 shows the results of estimating Equation 5.2 for prices, the number of bookings, the number of days booked, and revenues. $\beta_1$ is statistically different from zero at the threshold between 4 and 4.5 for variables capturing quantities and the revenue. There is no statistically significant effect for prices at this threshold.

At the threshold between 4.5 and 5 stars, $\beta_1$ is significant (although only marginally) for all variables. It is noteworthy that the effect of the transition between 4.5 and 5 stars is generally less pronounced than for the transition between 4 and 4.5 stars. By and large, the results validate the findings from the fixed effect analysis according to which rating positively affect revenues through increases in occupancy. Note that both approaches also deliver similar results in terms of effect-magnitudes.

## 5.5 The Impact of Prices on Ratings

In this section, we regress ratings on prices to analyze the extent to which prices might affect ratings. For each rating category, the overall rating and the six subcategories, we run the following regression:

$$rat_{it} = \beta_0 + \beta_1 ln(pr_{it-1}) + \tau_1 rc_{it} + \tau_2 rc_{it}^2 + \mu_i + \lambda_t + \epsilon_{it} \tag{5.3}$$

The review count ($rc$) enters as second order polynomial to control for the reduced sensitivity of the aggregate rating to new incoming ratings as the number of reviews increases. The coefficient $\beta_1$ in Equation 5.3 captures the impact of a one percent price increase on the aggregate rating. $\mu_i$ and $\lambda_t$ denote listing and month fixed effects, respectively. Standard errors are clustered at the listing level.

**Table 5.2:** Impact of prices on rating categories

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| | ov | ov-salient | value | comm | location | accuracy | checkin | clean |
| price | -0.0455*** | -0.0329*** | -0.0585*** | -0.0202* | -0.00457 | -0.0400*** | -0.0178* | -0.0153 |
| | (0.00754) | (0.00916) | (0.00966) | (0.00851) | (0.00724) | (0.00885) | (0.00834) | (0.00968) |
| Observations | 195582 | 195582 | 195582 | 195582 | 195582 | 195582 | 195582 | 195582 |

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5.2 shows the results from estimating Equation 5.3 for the various ratings. For the overall rating, we report the results for the salient and non-salient rating measure. The non-salient rating measure, which is more granular, is re-scaled to make coefficients comparable across specifications.

The magnitude of the price coefficient is highest for the value-for-money rating. For the rating subcategories other than value-for-money, prices have a smaller impact. For the overall rating, we find a negative and significant price coefficient. The results indicate that customers tend to factor negative prices in the overall rating. By contrast, prices appear to be less strongly correlated with ratings related to specific non-price aspects. According to the coefficients in Table 5.2, a ten percent price increase reduces the value-for-money rating by approximately 0.6 points, which corresponds to 0.3 stars.

In the following, we aim to decompose the aggregate negative coefficients documented in Table 5.2 by analyzing how prices correlate with the probability of observing a decrease or an increase in aggregate ratings. To do so, we estimate the following equation:

$$\mathbb{1}(rat_{it} < rat_{it-1}) = \beta_0 + \beta_1 ln(pr_{it-1}) + \tau_1 rc_{it} + \tau_2 rc_{it}^2 + \mu_i + \lambda_t + \epsilon_{it} \tag{5.4}$$

$\mathbb{1}$ is an indicator variable that takes the value one if the rating observed in $t$ is strictly smaller than the rating observed in $t-1$. In an analogous manner, we investigate the impact of price changes on the probability of observing an increase in aggregate ratings (replace $\mathbb{1}(rat_{it} < rat_{it-1})$ by $\mathbb{1}(rat_{it} > rat_{it-1})$).

**Table 5.3:** Impact of prices on probability of observing a negative rating change

| | (1)<br>ov | (2)<br>ov-salient | (3)<br>value | (4)<br>comm | (5)<br>location | (6)<br>accuracy | (7)<br>checkin | (8)<br>clean |
|---|---|---|---|---|---|---|---|---|
| price | 0.0521*** | 0.0120* | 0.0221*** | 0.0063 | -0.0075 | 0.0138* | -0.0002 | 0.0019 |
| | (0.0112) | (0.0061) | (0.0066) | (0.0051) | (0.0047) | (0.0059) | (0.0052) | (0.0065) |
| Observations | 147461 | 147461 | 147461 | 147461 | 147461 | 147461 | 147461 | 147461 |

Standard errors in parentheses
$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Table 5.3 reports the results from estimating Equation 5.4. The price coefficient for the non-salient overall rating is higher than for the salient rating. This is natural because the non-salient measure is more granular and, therefore, has a higher probability to change. According to the coefficients in Table 5.3, a one percent price increase increases the probability of a rating depreciation by five percent for the overall non-salient aggregate rating. Of the rating subcategories, the price coefficients is largest for the value-for-money.

**Table 5.4:** Impact of prices on probability of observing a positive rating change

| | (1)<br>ov | (2)<br>ov-salient | (3)<br>value | (4)<br>comm | (5)<br>location | (6)<br>accuracy | (7)<br>checkin | (8)<br>clean |
|---|---|---|---|---|---|---|---|---|
| price | -0.0807*** | -0.0155* | -0.0389*** | -0.0201*** | -0.0117* | -0.0215*** | -0.0105* | -0.0200*** |
| | (0.0117) | (0.0063) | (0.0061) | (0.0052) | (0.0054) | (0.0058) | (0.0053) | (0.0061) |
| Observations | 147461 | 147461 | 147461 | 147461 | 147461 | 147461 | 147461 | 147461 |

Standard errors in parentheses
$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Table 5.4 shows the analogous results obtained when regressing the indicator variable for observing an increase in ratings on prices. According to the results, higher prices negatively correlate with the probability of aggregate rating increases across all categories. Again, we find that, for the salient measures, the coefficient is the largest for the value-for-money rating.

The results of the analysis for rating changes in Tables 5.3 and 5.4 are consistent with the results in Table 5.2. Prices are positively correlated with a decrease in rating *and* negatively correlated with an increase in rating only for the value-for-money and overall ratings. For the other specific ratings categories, prices are only negatively correlate with an increase in rating (except for the accuracy rating). Altogether, the results from this Section provide empirical support for the hypothesis that consumers tend to factor prices into their overall assessment of the service.[11]

## 5.6 The Impact of Low Entry Prices on the Long-Term Performance of Listings

The results of Section 5.4 reveal a clear positive relationship between ratings and the performance of a listing. Obviously, there is a clear incentive for hosts to improve ratings in order to boost revenues. The results of Section 5.5 provide evidence that prices negatively correlate with the ratings of a listing.

This Section analyzes whether the data back the hypothesis that strategic price-setting might influence the long-term performance of a listing. We specifically focus on the initial prices set by hosts when first entering the market. From the results of Section 5.5, we would expect that, ceteris paribus , listings with lower initial prices achieve higher initial ratings. Our results confirm this conjecture. Additionally, we find evidence that listings with low initial entry prices are subsequently able to increase prices: While this leads to a depreciation of ratings in accordance with the results from Section 5.5, these listings are nevertheless able to maintain a rating advantage compared to listings that initially charged higher prices. A possible explanation for this finding is that the initial rating advantage obtained by the initially lower prices caries over through behavioral biases: Consumers might tend to orient themselves on the aggregate ratings they observe when submitting their own ratings.

The empirical strategy is akin to a difference in differences design. We compare the performance of listings charging a low entry-price with listings charging a high entry price. To do so, we create a variable that measures the entry price relative to the average price charged in all subsequent periods. For the analysis, we focus only on listings with less than three reviews the first time we observe them in the data. Additionally, we impose the restriction that a listing must be observed a least five times to track the evolution of performance measures and ratings over a meaningful time span. We run the following regression:

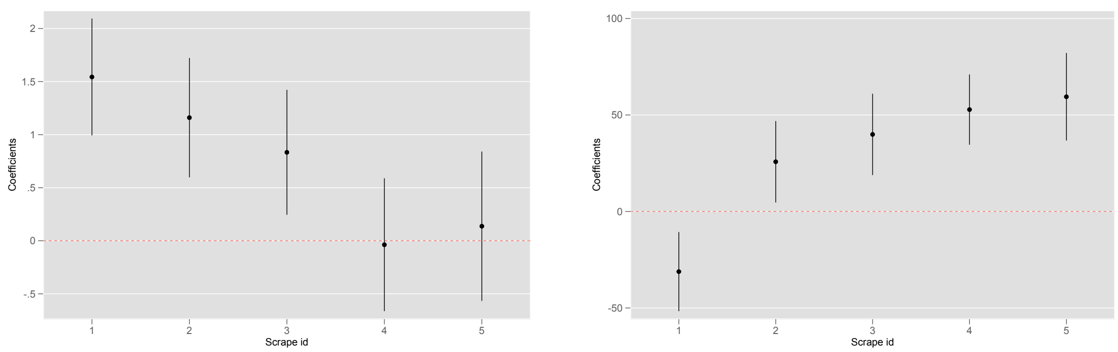
$$y_{i\tau} = \beta_0 + \sum_{j=1}^{j=5} \beta_\tau \times fpr_i \times \mathbb{1}[\tau = j] + X_{i\tau} + loc_i + \epsilon_{i\tau} \tag{5.5}$$

$fpr_i$ is defined by $pr_{i\tau=1}/\overline{pr}_{i\tau>1}$. $p_{i\tau=1}$ is the initial price charged by a listing. $\overline{pr}_{i\tau>1}$ denotes the average price charged in all subsequent periods. $\mathbb{1}[\tau = j]$ is a dummy variable for the $jth$ time a listing is observed. The coefficient $\beta_\tau$ captures the impact of $fpr_i$ on the outcome in period $\tau$. $X_{i\tau}$ contains time-varying listing specific characteristics, including the absolute average price, the

[11] In Appendix 5.8.4, we conduct a series of robustness check. We show that we obtain qualitatively similar results when focusing only on observations for which we observe fewer than 20 reviews, which allows to closer approximate the impact of prices on individual ratings. We further show that the results focusing on rating changes are robust when using a logit model instead of a linear probability model.

number of reviews and the number of days a listing was available in a given month. $loc_i$ denotes a set of dummy variables that control for the location of a listing.[12] Note that $\tau$ denotes the number of times a listing has been observed. To account for seasonal effects, we also control for month fixed effects. We further include a set of dummy variables that control for the first month in which a listing has been observed. We omit these controls from the notation in Equation 5.5 for expositional reasons (i.e. we reverse the sign of the coefficients in the above examples).

For illustrative purposes, consider the following examples: $fpr_i = 0.8$ means that the entry price was 20 percentage points lower than the average price charged by a listing for $\tau > 1$. If $y_{i\tau}$, the outcome of interest, is the booking count and $\beta_1 = -1$, this means that increasing the relative entry price by 100 percent reduces the number of bookings by 1 for in the first period. If $\beta_{\tau=2} = -0.4$, this means that the number of bookings in the second period is reduced by 0.4. Thus, the specification in Equation 5.5 allows for assessing how differences in the relative entry price affect listing performance over time. To facilitate interpretation of the results, we report the results in terms of an impact of a *reduction* of the relative entry price.




**(a)** $\beta_\tau$ for $y_{i\tau}$ = booking count per month **(b)** $\beta_\tau$ for $y_{i\tau}$ = price

**Figure 5.3:** Partial Correlation of $fpr_i$ with bookings and prices.

Figure 5.3 displays the $\beta_\tau$ regression coefficients from Equation 5.5 when considering bookings (left panel) and prices (right panel) as the dependent variable. From the left panel of Figure 5.3, we see that listings with a lower relative entry price are booked more often in the first period (i.e. when they enter the market). This surplus in bookings vanishes after four periods. From the right panel of Figure 5.3, we see that the steady decline in bookings is accompanied by a steady increase in prices. Initially, listings with a low initial relative entry price charge lower prices, as seen in the negative coefficient for $\tau = 1$. The sign of the difference reverses after the first period. This means that, after an initial period of very low prices, listings with a low initial entry price tend to set higher prices compared to listings with high initial entry prices.

Figure 5.4 reveals what the combination of increasing prices and decreasing bookings implies in terms of revenue. In the first period, listings with a low initial relative entry price achieve the same revenue as listings with a high initial entry price: The negative revenue impact of lower prices is fully compensated by the additional bookings these lower prices generate. In subsequent

[12]We can geolocate the listings with a precision of approximately 250 meters. This allows us to determine the statistical area a listing is located in. The dummies we use capture the most granular statistical area in France. For the city of Paris there are 800 such areas, the average size is slightly less than 0.13 square-kilometer. Thus, the dummies capture very granular geographic heterogeneity.

periods, the increase in prices overcompensates the reduction in bookings such that listings with lower relative entry prices achieve higher revenue than listings with higher relative entry prices.

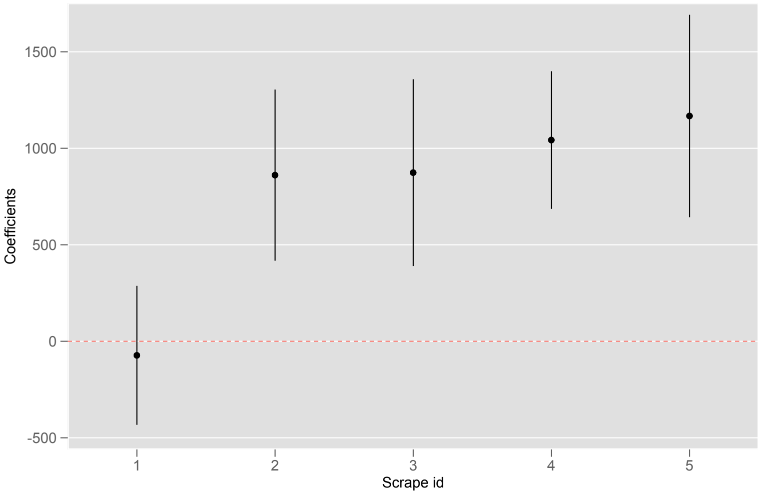
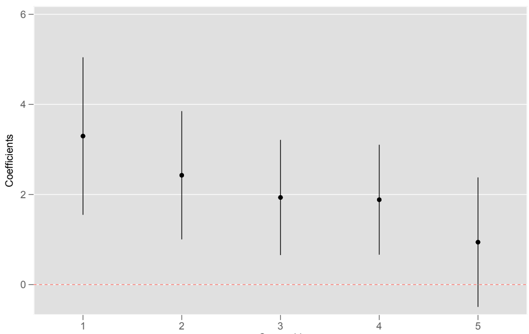
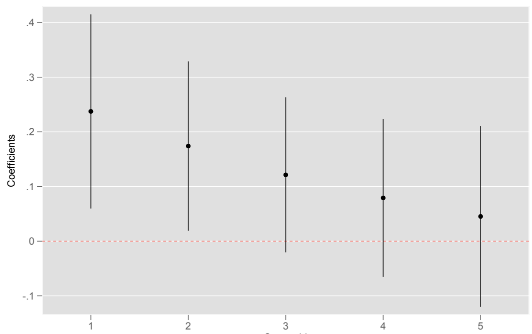



**Figure 5.4:** Partial Correlation of $fpr_i$ with revenues.

Figure 5.5 summarizes the evolution of the overall rating and the value-for-money rating as a function of $fpr_i$. Consistent with our previous analysis, listings with a low initial entry price have a higher overall and value-for-money rating in the initial period. Also consistent with our previous results, the rating advantage for these two rating categories decreases as listings with a low initial price raise their prices. Note, however, that despite charging higher prices in periods starting from $\tau = 2$, the listings with a low initial entry price do not have worse ratings in these two categories than listings that charged high initial entry prices. Thus, the initial advantage seems to carry over into subsequent periods. Interestingly, we find no indication of significant differences between listings with high and low entry prices for all other rating subcategories except the accuracy rating (which is consistent with the results from Section 5.5 in which we found that the accuracy rating appears to strongly correlate with prices). We present the corresponding results in Appendix 5.8.5.




**(a)** $\beta_\tau$ with $y_{i\tau}$ = overall rating




**(b)** $\beta_\tau$ with $y_{i\tau}$ = value-for-money rating

**Figure 5.5:** Partial Correlation of $fpr_i$ with the overall and value-for-money rating.

The results of this Section are broadly consistent with hosts gaining a long-run advantage in ratings when charging low initial prices. The evolution of the overall and value-for-money ratings provided additional evidence that higher prices depreciate ratings. Unobserved factors, like the time-constant quality of an individual listing, are not accounted for in the present analysis. Since a fixed effect approach would erase level-effects, it would not allow us, for instance, to identify the impact of relative entry prices in $\tau = 1$. For instance, with fixed effects, we would only be able to identify the declining trend in Figure 5.5. In Appendix 5.8.5, we show that with or without fixed effect, the measured magnitude of the declining trend would be almost identical. This is comforting as it suggests that the omission of fixed effects does not drive our results.

# 5.7 Conclusion

This paper provides additional evidence to the existing empirical literature finding a positive relationship between ratings and revenues of sellers on online platforms. The paper further analyzes the impact of prices on ratings. The findings provide evidence that higher prices depreciate ratings. The results provide empirical support for the hypothesis that prices tend to negatively affect the overall satisfaction with an accommodation despite guests having the option to rate the value-for-money aspect of their trip in a separate rating category.

Prices correlate the most with the overall rating and the value-for-money rating. The results suggest that non-price rating categories are only affected to a lesser extend. For the non-price rating categories, higher prices only seem to reduce the probability of an increase in ratings. By contrast, higher prices seem to increase the probability of rating decreases only for the overall and value-for-money categories. The fact that the overall rating has a very similar price dependence to the value-for-money rating is yet another piece of evidence that customers tend to assess quality net of price rather than just quality.

The paper concludes with an analysis of the rating and performance dynamics. This analysis highlights the apparent importance of entry-prices and initial ratings. We find evidence that low entry prices tend to increase initial ratings and that hosts can leverage this higher initial ratings to charge higher prices in the long-run. This analysis suffers from the natural limitations of observational data and needs to be confirmed by additional experimental evidence. It also offers interesting avenues for further research regarding how behavioral biases might perpetuate high ratings despite decreasing quality.

## 5.8 Appendix

### 5.8.1 Leaving a Review on Airbnb

The following series of screenshots summarizes step-by-step the review procedure on Airbnb. The review was performed by the author in September 2020. The names of the accommodation and host are anonymized. Guests in 2017, the time span of the analysis, likely experienced a slightly different review process. Nevertheless, the screenshots highlight features that likely did not change much over time, including the fact that the overall rating is collected before the ratings on the specific subcategories.

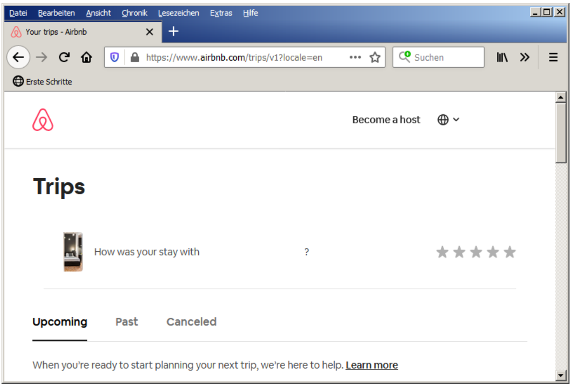



**Figure 5.6:** [step 1] After log-in into the Airbnb account, the guest is asked " how was your stay?" On the right, the guest can leave a star rating from one to five (only full stars are allowed). The star rating assigned here corresponds to the overall rating.

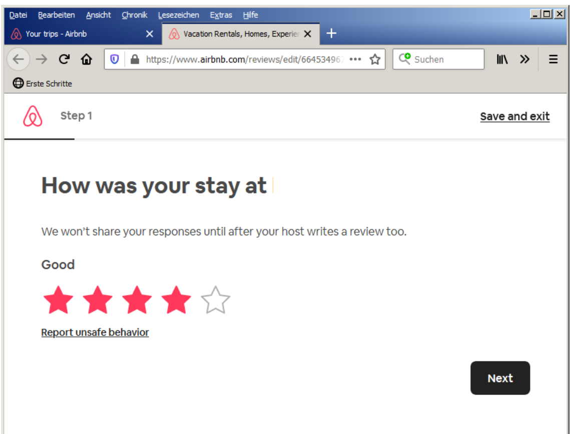



**Figure 5.7:** [step 2] After submitting the overall rating in [step 1], the guest is asked to confirm the rating. Airbnb informs the host that the rating will only be " shared" after the host rated the guest.

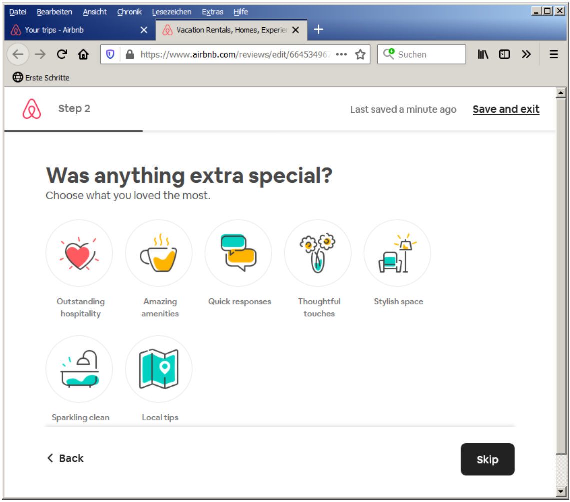



**Figure 5.8:** [Step 3] The guest is prompted to highlight special positive features of his stay. This is optional as can be seen from the button "skip" on the bottom right.

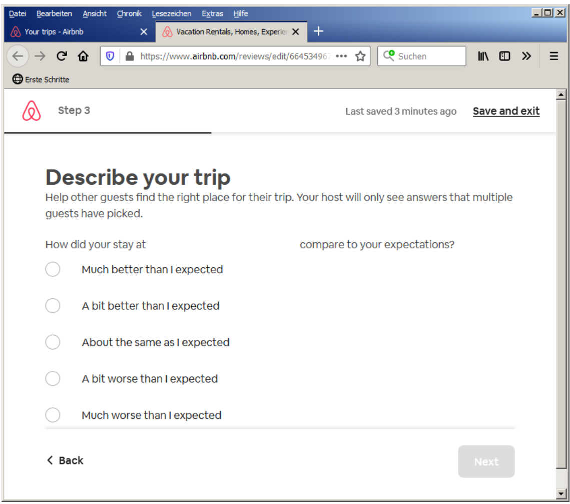



**Figure 5.9:** [Step 4] The guest is asked to benchmark his experience relative to his expectations. This step is not optional.

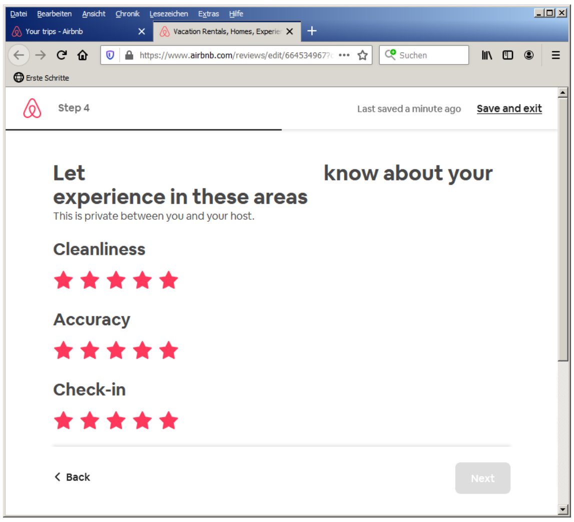



**Figure 5.10:** [Step 5] The guest is asked to rate the accommodation/ stay with respect to the subcategories described in the main text (only full stars are allowed).

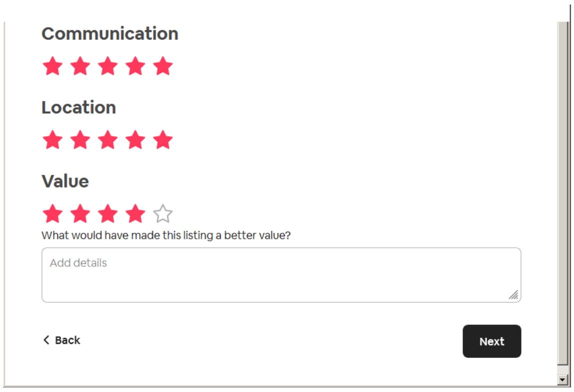



**Figure 5.11:** [Step 5 contd.] Note that for any rating of less than 5 stars, the guest is given the option to explain his choice.

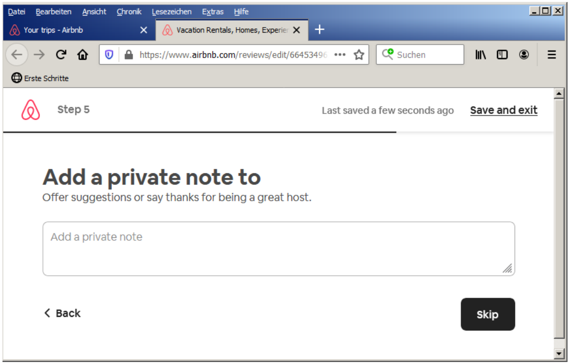



**Figure 5.12:** [Step 6] The guest is given the possibility to leave a private note to the host.

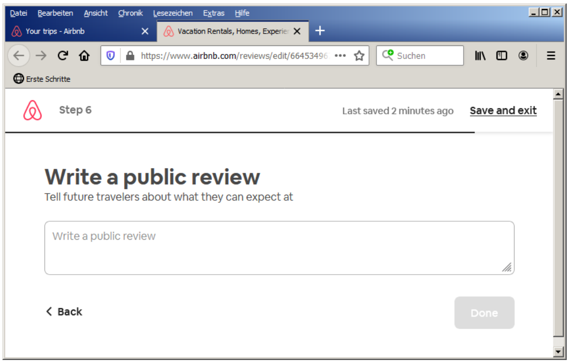



**Figure 5.13:** [Step 7] The guest is asked to leave a public review. This is the final step of the review process.

### 5.8.2 Aggregation of individual ratings

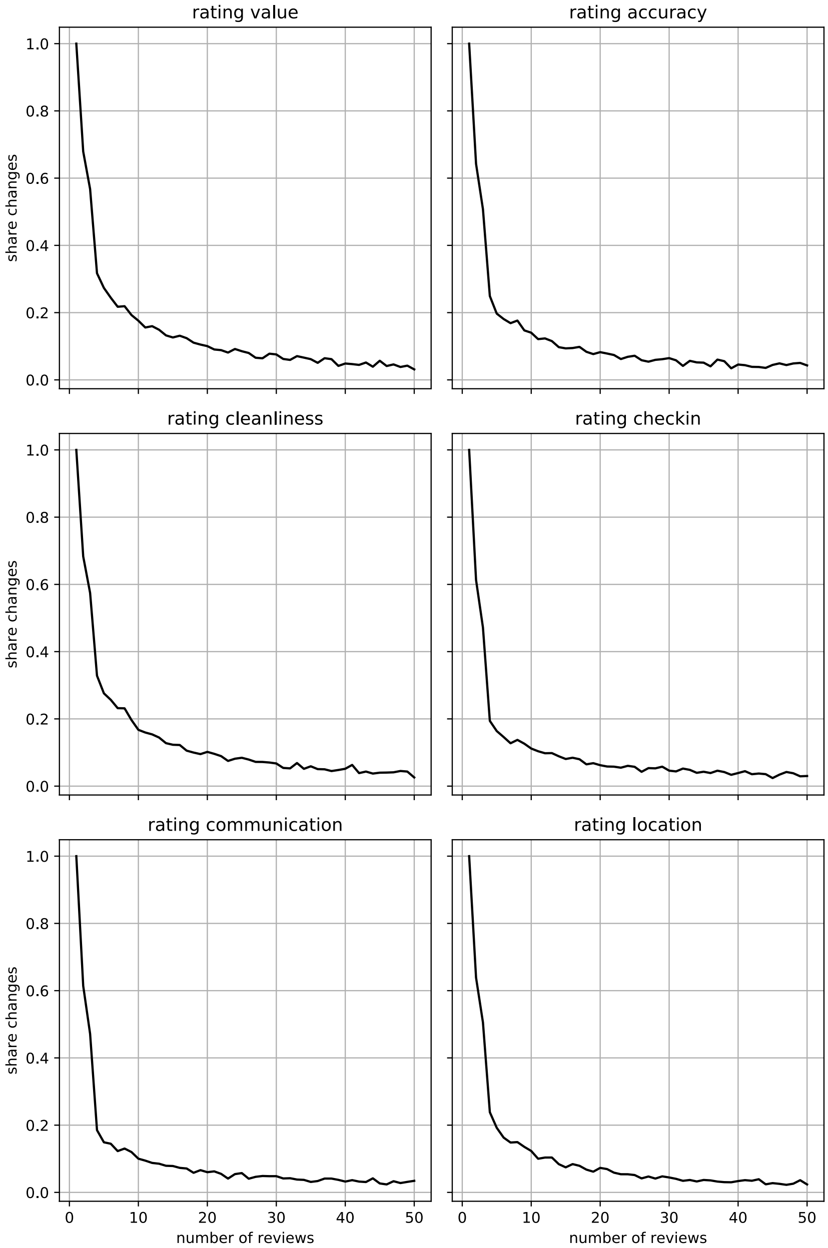



**Figure 5.14:** Relative frequency of rating changes for the different rating subcategories

### 5.8.3 The Impact of Ratings on Performance

In this appendix, we present robustness results of the analysis in Section 5.4. Table 5.5 includes the ratings of the other subcategories. The other rating subcategories do not seem to have a consistent impact on the various performance measures. Note that, occasionally, some ratings subcategories seem to have a larger impact than the overall rating. Compared to the regression specification in Section 5.4, the impact of the overall star rating decreases. Nevertheless, the main conclusion that we drew from Section 5.4 hold.

**Table 5.5:** Impact of overall star rating on several performance measures

| | (1)<br>price | (2)<br>bookings | (3)<br>days booked | (4)<br>revenue |
|---|---|---|---|---|
| rating: overall | 0.142 | 0.0487* | 0.293*** | 24.64*** |
| | (0.165) | (0.0212) | (0.0490) | (6.977) |
| rating: value | 0.133 | 0.0390 | 0.0384 | -3.072 |
| | (0.177) | (0.0199) | (0.0464) | (7.114) |
| rating: comm. | 0.0181 | 0.0823** | 0.0879 | 27.44** |
| | (0.212) | (0.0255) | (0.0579) | (9.699) |
| rating: location | -0.167 | 0.0503* | 0.118* | 7.882 |
| | (0.175) | (0.0229) | (0.0515) | (7.268) |
| rating: clean | 0.239 | 0.0347 | 0.158*** | 15.36* |
| | (0.147) | (0.0185) | (0.0438) | (6.115) |
| rating: accuracy | 0.0429 | 0.0186 | 0.0382 | 6.157 |
| | (0.173) | (0.0221) | (0.0501) | (7.599) |
| rating: check-in | 0.0116 | -0.00188 | 0.0235 | 8.958 |
| | (0.236) | (0.0238) | (0.0543) | (8.625) |
| days available | 0.0145 | 0.150*** | 0.813*** | 71.12*** |
| | (0.00849) | (0.000998) | (0.00235) | (0.499) |
| Constant | 85.67*** | -2.181*** | -11.89*** | -1296.1*** |
| | (2.596) | (0.327) | (0.742) | (109.9) |
| Observations | 169880 | 169880 | 169880 | 169880 |

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Compared to Table 5.5, Table 5.6 additionally includes measures capturing the demand state. "Occupancy rivals" denotes the fraction of booked listings in the neighborhood of a listing. "Offer rivals" denotes the number of listings available for booking in the neighborhood of a listing. The neighborhood is defined by the smallest statistical area in France, of which they are 800 in the city of Paris. Additionally, the regression specification in Table 5.6 controls for the number of reviews and the price of a listing, where appropriate. As shown, the main conclusions drawn in Section 5.4 remain unaffected.

**Table 5.6:** Impact of overall star rating on several performance measures

| | (1)<br>price | (2)<br>bookings | (3)<br>days booked | (4)<br>revenue |
|---|---|---|---|---|
| rating: overall | 0.145 | $0.0426^{*}$ | $0.273^{***}$ | $20.96^{**}$ |
| | (0.165) | (0.0212) | (0.0486) | (6.615) |
| rating: value | 0.130 | 0.0389 | 0.0394 | -5.354 |
| | (0.177) | (0.0199) | (0.0462) | (6.949) |
| rating: comm. | 0.0192 | $0.0839^{**}$ | 0.0930 | $27.46^{**}$ |
| | (0.211) | (0.0256) | (0.0574) | (8.986) |
| rating: location | -0.170 | $0.0476^{*}$ | $0.110^{*}$ | 10.27 |
| | (0.175) | (0.0229) | (0.0513) | (6.911) |
| rating: clean | 0.244 | 0.0344 | $0.155^{***}$ | 11.02 |
| | (0.147) | (0.0185) | (0.0435) | (6.060) |
| rating: accuracy | 0.0398 | 0.0199 | 0.0419 | 5.577 |
| | (0.173) | (0.0221) | (0.0497) | (7.391) |
| rating: check-in | 0.00847 | -0.00205 | 0.0218 | 8.584 |
| | (0.236) | (0.0238) | (0.0539) | (8.200) |
| number reviews | 0.00551 | $-0.0152^{***}$ | $-0.0474^{***}$ | $-2.780^{***}$ |
| | (0.00876) | (0.00125) | (0.00205) | (0.273) |
| days available | 0.0157 | $0.150^{***}$ | $0.812^{***}$ | $70.85^{***}$ |
| | (0.00847) | (0.00100) | (0.00236) | (0.490) |
| occupancy rivals | $12.29^{***}$ | -0.0731 | $-2.619^{***}$ | -60.60 |
| | (0.848) | (0.101) | (0.250) | (41.12) |
| offer rivals | $0.0206^{**}$ | $0.00253^{***}$ | $0.0180^{***}$ | $1.638^{***}$ |
| | (0.00786) | (0.000748) | (0.00182) | (0.270) |
| price | | $-0.00681^{***}$ | $-0.0145^{***}$ | $16.80^{***}$ |
| | | (0.000471) | (0.00129) | (0.746) |
| Constant | $80.32^{***}$ | $-1.401^{***}$ | $-10.46^{***}$ | $-2806.9^{***}$ |
| | (2.717) | (0.341) | (0.774) | (117.8) |
| Observations | 169880 | 169880 | 169880 | 169880 |

Standard errors in parentheses

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

### 5.8.4 The Impact of Prices on Ratings

Table 5.7 shows the impact of prices on the various rating categories for observations with less than 20 reviews. As we elaborated in Section 5.3, for listings with fewer reviews, individual ratings have a stronger impact on the aggregate ratings. Again, we find that prices have the strongest impact on the value rating, which appears natural. The overall rating also seems strongly affected. For the other specific rating categories, we also find a negative impact, which is always smaller in magnitude than for the value-for-money rating. Like with the results in Section 5.5 , we find a strong impact on the accuracy rating, which is the only one of the specific rating subcategories for which the price coefficient is comparable to the price coefficient for the overall rating.

Table 5.8 shows the impact of prices on the probability of observing a rating depreciation when only including observations with less than 20 reviews. Only the value rating and the granular overall rating measure react significantly to price changes. The same is true when analyzing the impact of prices on the probability of rating appreciations.

Generally, when only utilizing observations with less than 20 Reviews, the results are broadly in line with the results of the main section.

**Table 5.7:** Impact of prices on rating categories

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| | ov | ov-salient | value | comm | location | accuracy | checkin | clean |
| price | -0.0683*** | -0.0599*** | -0.108*** | -0.0439** | -0.00846 | -0.0697*** | -0.0309* | -0.0363* |
| | (0.0146) | (0.0167) | (0.0176) | (0.0152) | (0.0131) | (0.0159) | (0.0146) | (0.0179) |
| Observations | 100152 | 100152 | 100152 | 100152 | 100152 | 100152 | 100152 | 100152 |

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

**Table 5.8:** Impact of prices on probability of observing a negative rating change

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| | ov | ov-salient | value | comm | location | accuracy | checkin | clean |
| price | 0.0475* | 0.0180 | 0.0413** | 0.0130 | -0.0151 | 0.0129 | -0.0010 | -0.0036 |
| | (0.0208) | (0.0134) | (0.0145) | (0.0114) | (0.0110) | (0.0131) | (0.0114) | (0.0148) |
| Observations | 62444 | 62444 | 62444 | 62444 | 62444 | 62444 | 62444 | 62444 |

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

**Table 5.9:** Impact of prices on probability of observing a positive rating change

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| | ov | ov-salient | value | comm | location | accuracy | checkin | clean |
| price | -0.0436* | -0.0081 | -0.0473*** | -0.0163 | -0.0144 | -0.0102 | -0.0136 | -0.0200 |
| | (0.0202) | (0.0135) | (0.0138) | (0.0108) | (0.0122) | (0.0124) | (0.0116) | (0.0132) |
| Observations | 62444 | 62444 | 62444 | 62444 | 62444 | 62444 | 62444 | 62444 |

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Tables 5.10 and 5.11 show the results of using a logit specification with fixed effects to analyze the probability of rating changes. Note that to estimate a logit specification with fixed effects, one must drop listings for which no change in the dependent variable is observed.

**Table 5.10:** Impact of prices on probability of observing a negative rating change

| | (1) ov | (2) ov-salient | (3) value | (4) comm | (5) location | (6) accuracy | (7) checkin | (8) clean |
|---|---|---|---|---|---|---|---|---|
| main | | | | | | | | |
| price | 0.3295*** | 0.2607 | 0.4057*** | 0.1787 | -0.1514 | 0.3238* | 0.0036 | 0.0524 |
| | (0.0725) | (0.1331) | (0.1227) | (0.1544) | (0.1733) | (0.1405) | (0.1475) | (0.1239) |
| Observations | 90956 | 29264 | 33847 | 20652 | 18329 | 25183 | 21271 | 33255 |

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

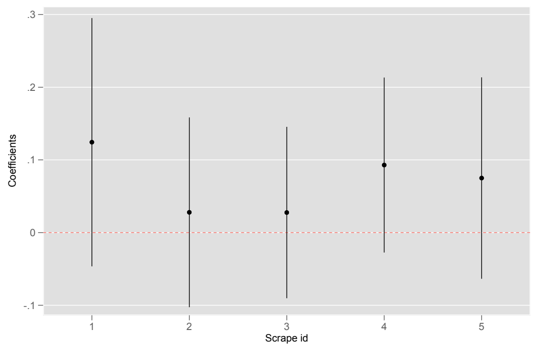
**Table 5.11:** Impact of prices on probability of observing a positive rating change

| | (1) ov | (2) ov-salient | (3) value | (4) comm | (5) location | (6) accuracy | (7) checkin | (8) clean |
|---|---|---|---|---|---|---|---|---|
| main | | | | | | | | |
| price | -0.4755*** | -0.2919* | -0.8098*** | -0.5814*** | -0.2435 | -0.5075*** | -0.2496 | -0.3900** |
| | (0.0696) | (0.1339) | (0.1378) | (0.1591) | (0.1550) | (0.1444) | (0.1512) | (0.1413) |
| Observations | 101953 | 30265 | 30793 | 19899 | 23530 | 27270 | 22157 | 27901 |

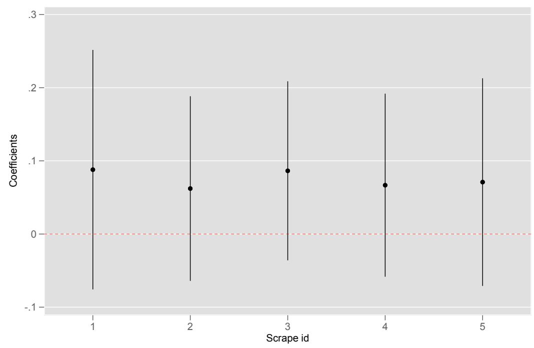
Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

### 5.8.5 The Impact of Low Entry Prices on the Long-Term Performance of Listings

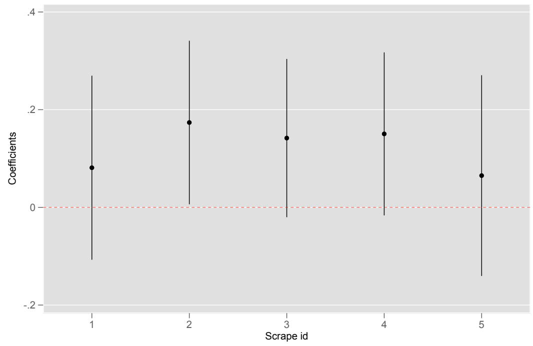
Figure 5.15 shows the partial correlation between $fpr_i$ and the various specific rating subcategories. As mentioned in the main text, the coefficients for the accuracy rating follow a very similar pattern to the coefficients for the overall and value-for-money rating. This is consistent with the results from Section 5.5. Figures 5.16, 5.17 and 5.18 compare the $\beta_{\tau>2}$ coefficients between a fixed effect regression and the OLS specification used in the main text. As shown, the potential bias from omitting fixed effects seems small and is likely not driving the main results.



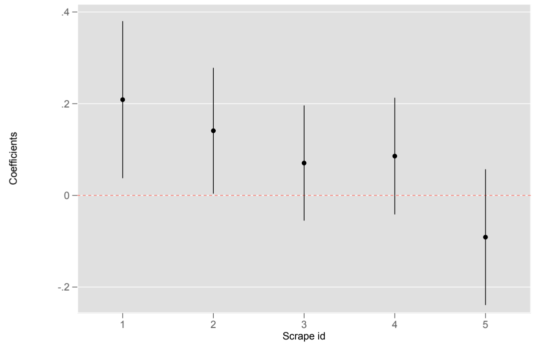

**(a)** $\beta_\tau$ for $y_{i\tau}$ = communication rating



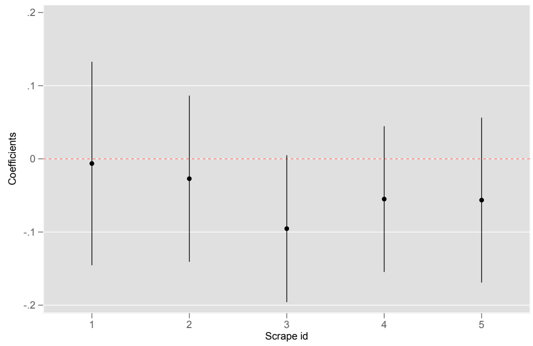

**(b)** $\beta_\tau$ for $y_{i\tau}$ = check-in rating




**(c)** $\beta_\tau$ for $y_{i\tau}$ = cleanliness rating

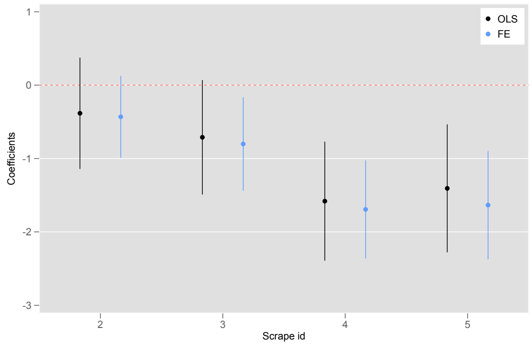



**(d)** $\beta_\tau$ for $y_{i\tau}$ = accuracy rating

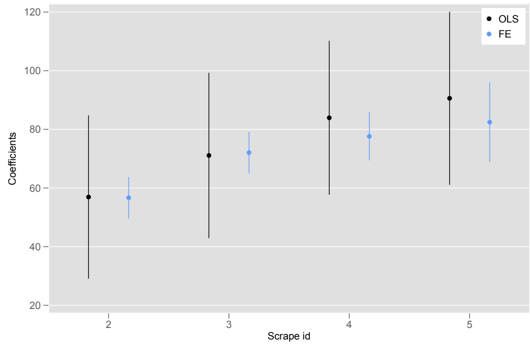



**(e)** $\beta_\tau$ for $y_{i\tau}$ = location rating

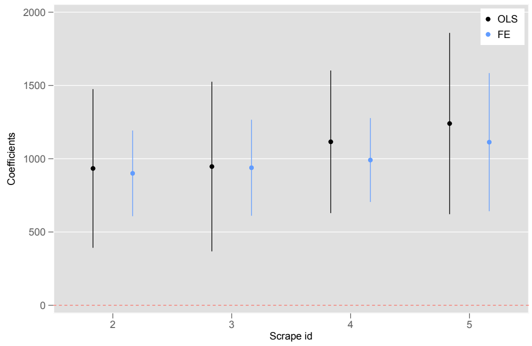
**Figure 5.15:** Partial Correlation of $fpr_i$ with different rating subcategories




**(a)** Monthly booking count




**(b)** Price




**(c)** Revenue

**Figure 5.16:** Comparison of $\beta_{\tau>2}$ between OLS and fixed-effect for different performance measures

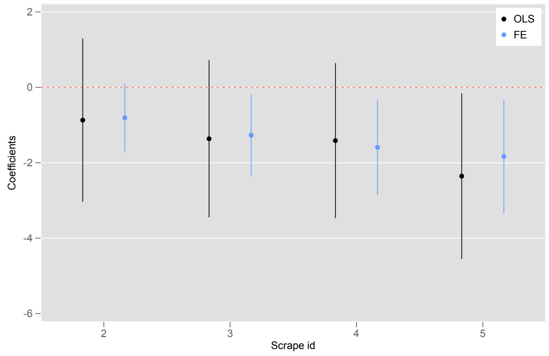
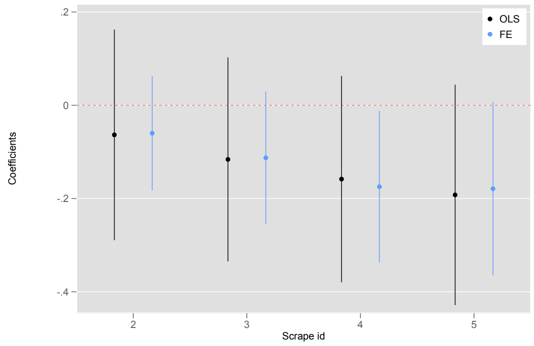
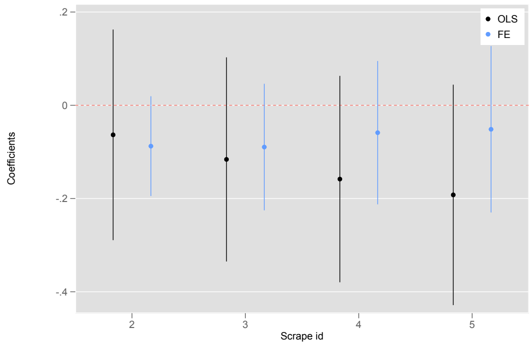


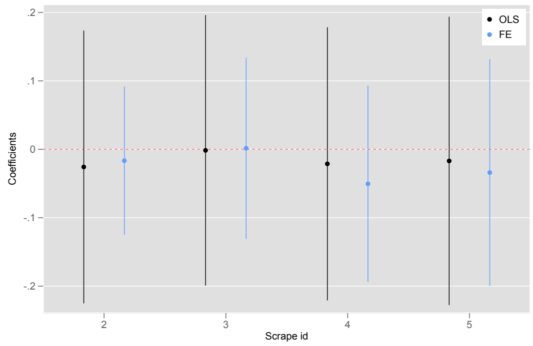

**Figure 5.17:** Comparison of $\beta_{\tau>2}$ between OLS and fixed-effect for overall rating



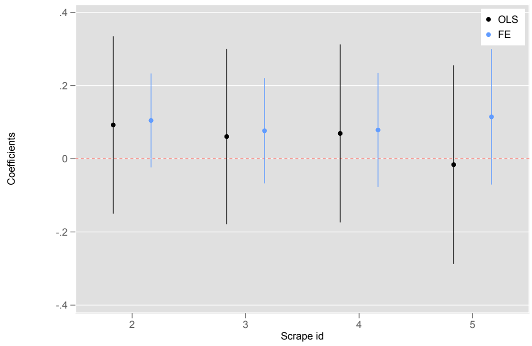

**(a)** Value rating



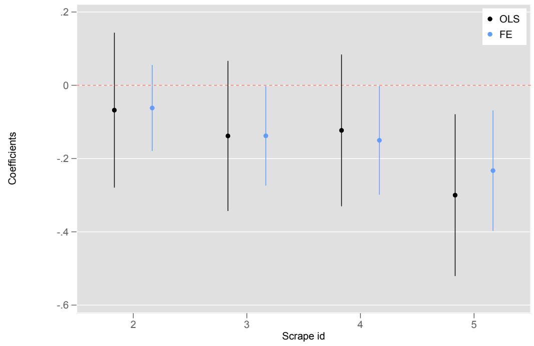

**(b)** Communication rating



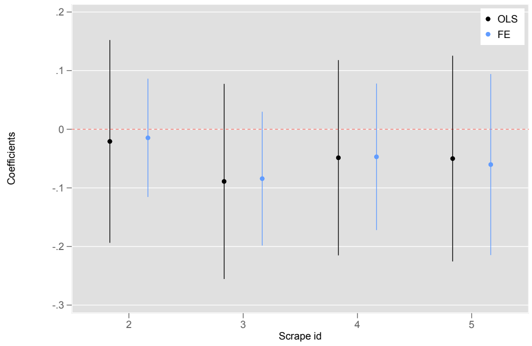

**(c)** Check-in rating




**(d)** Cleanliness rating




**(e)** Accuracy rating




**(f)** Location rating

**Figure 5.18:** Comparison of $\beta_{\tau>2}$ between OLS and fixed-effect for different rating subcategories

# Chapter 6

# Conclusion

The rise of digitization transforms our economy. This dissertation attempts to provide scientific insights to inform the policy debate on how to best tackle some of the challenges related to the digital transformation.

The second chapter of the present dissertation addresses the increasingly relevant question about the role of data for market power. The analysis uses search engine data from Yahoo! from the year 2010. The article discusses a potential mechanism rationalizing the so-called “data-feedback loop” hypothesis in the absence of increasing returns to scale. The empirical analysis provides evidence in favor of the hypothesized mechanism.

The third and fourth chapter of this dissertation address economic externalities caused by the rise of the platform economy. The third chapter focuses on the welfare impact of Airbnb on the hospitality sector. Using a discrete choice model that allows for realistic demand features, we quantify the positive impact of Airbnb on travelers. Additionally, we quantify by how much Airbnb reduces revenues of hotels by calibrating a flexible supply side model. Our results suggest that the net impact of Airbnb on welfare in the hospitality market is positive. The fourth chapter assesses the impact of Airbnb on housing. Using plausibly exogenous variation in the supply of Airbnb, we quantify the impact of Airbnb on asked rents in Berlin. Additionally, we provide new insights with respect to effect heterogeneity by showing that high availability Airbnbs appear to be driving rent increases the most.

The fifth chapter of the dissertation sheds light on how factors unrelated to quality can influence ratings. With the increasing importance of online market places, questions about the role of ratings for the success of sellers and how rating systems should be designed have attracted the attention of economic researchers. The chapters reveals an interesting negative correlation between prices and ratings. Additionally, the chapter provides suggestive evidence on how sellers might strategically exploit price setting to boost their revenues through higher ratings.

# Bibliography


Ackerberg, D. A. and Rysman, M. (2005) Unobserved Product Differentiation in Discrete-Choice Models: Estimating Price Elasticities and Welfare Effects, *The RAND Journal of Economics*, **36**, 771–788.

Adomavicius, G. and Tuzhilin, A. (2005) Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions, *IEEE Transactions on Knowledge & Data Engineering*, **17**, 734–749.

Aguiar, L. and Waldfogel, J. (2018) As streaming reaches flood stage, does it stimulate or depress music sales?, *International Journal of Industrial Organization*, **57**, 278–307.

Argenton, C. and Prüfer, J. (2012) Search Engine Competition with Network Externalities, *Journal of Competition Law and Economics*, **8**, 73–105.

Bajari, P., Chernozhukov, V., Hortaçsu, A. and Suzuki, J. (2019) The Impact of Big Data on Firm Performance: An Empirical Investigation, in *AEA Papers and Proceedings*, vol. 109, pp. 33–37.

Barnett, A. G., van der Pols, J. C. and Dobson, A. J. (2004) Regression to the Mean: What It Is and How to Deal With It, *International Journal of Epidemiology*, **34**, 215–220.

Barron, K., Kung, E. and Proserpio, D. (2020) The Effect of Home-Sharing on House Prices and Rents: Evidence from Airbnb, *Marketing Science*, **forthcoming**.

Basuroy, S., Kim, Y. and Proserpio, D. (2020) Estimating the impact of Airbnb on the local economy: Evidence from the restaurant industry, *Working Paper*.

Belloni, A., Chen, D., Chernozhukov, V. and Hansen, C. (2012) Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain, *Econometrica*, **80**, 2369–2429.

Belloni, A., Chernozhukov, V. and Hansen, C. (2014) Inference on Treatment Effects after Selection among High-Dimensional Controls, *The Review of Economic Studies*, **81**, 608–650.

Berry, S., Levinsohn, J. and Pakes, A. (1995) Automobile Prices in Market Equilibrium, *Econometrica*, **63**, 841–890.

Berry, S. T. (1994) Estimating discrete-choice models of product differentiation, *The RAND Journal of Economics*, **25**, 242–262.

Bolton, G., Greiner, B. and Ockenfels, A. (2013) Engineering trust: reciprocity in the production of reputation information, *Management science*, **59**, 265–285.

Cabral, L. and Li, L. (2015) A dollar for your thoughts: Feedback-conditional rebates on ebay, *Management Science*, **61**, 2052–2063.

Cachon, G. P., Daniels, K. M. and Lobel, R. (2017) The Role of Surge Pricing on a Service Platform with Self-Scheduling Capacity, *Manufacturing & Service Operations Management*, **19**, 368–384.

Casadesus-Masanell, R. and Hervas-Drane, A. (2015) Competing with privacy, *Management Science*, **61**, 229–246.

Castillo, J. C. (2019) Who Benefits from Surge Pricing?, *SSRN*, `https://ssrn.com/abstract=3245533` (December 28, 2019).

Chernozhukov, V., Hansen, C. and Spindler, M. (2015) Post-Selection and Post-Regularization Inference in Linear Models with Many Controls and Instruments, *American Economic Review: Papers & Proceedings*, **105**, 486–490.

Chiou, L. and Tucker, C. (2017) Search Engines and Data Retention: Implications for Privacy and Antitrust, *NBER Working Paper Series*, **23815**.

Chuklin, A., Serdyukov, P. and De Rijke, M. (2013) Click Model-Based Information Retrieval Metrics, in *Proceedings of the 36th international ACM SIGIR conference on research and development in information retrieval*, ACM, pp. 493–502.

Claussen, J., Peukert, C. and Sen, A. (2019) The Editor vs. the Algorithm: Economic Returns to Data and Externalities in Online News, *SSRN*, `https://ssrn.com/abstract=3399947` (November 3, 2019).

Cohen, P., Hahn, R., Hall, J., Levitt, S. and Metcalfe, R. (2016) Using Big Data to Estimate Consumer Surplus: The Case of Uber, *NBER Working Paper Series*, **22627**.

Cramer, J. and Krueger, A. B. (2016) Disruptive Change in the Taxi Business: The Case of Uber, *American Economic Review: Papers & Proceedings*, **106**, 177–182.

de Palma, A., Picard, N. and Waddell, P. (2007) Discrete choice models with capacity constraints: An empirical analysis of the housing market of the greater Paris region, *Journal of Urban Economics*, **62**, 204–230.

Decarolis, F., Goldmanis, M. and Penta, A. (2020) Marketing agencies and collusive bidding in online ad auctions, *Management Science*, **66**, 4359–4377.

Decarolis, F. and Rovigatti, G. (2019) From Mad Men to Maths Men: Concentration and Buyer Power in Online Advertising, *CEPR Discussion Paper*, **13897**.

Drahokoupil, J. and Fabo, B. (2016) The platform economy and the disruption of the employment relationship, *ETUI Research Paper-Policy Brief*, **5**.

Dudás, G., Vida, G., Kovalcsik, T. and Boros, L. (2017) A socio-economic analysis of airbnb in new york city, *Regional Statistics*, **7**, 135–151.

Duso, T., Michelsen, C., Schaefer, M. and Tran, K. D. (2020) Airbnb and Rents: Evidence from

Berlin, *DIW Berlin Discussion Paper*, **1890**.

Eaton, D. H. *et al.* (2002) Valuing information: Evidence from guitar auctions on ebay, *Murray, KY, Murray State University*, **28**.

Edelman, B. G. and Geradin, D. (2015) Efficiencies and regulatory shortcuts: How should we regulate companies like airbnb and uber, *Stan. Tech. L. Rev.*, **19**, 293.

Einav, L., Farronato, C. and Levin, J. (2016) Peer-to-peer markets, *Annual Review of Economics*, **8**, 615–635.

European Commission (2018) Commission Decision AT.40099 – Google Android, `https://ec.europa.eu/competition/antitrust/cases/dec_docs/40099/40099_9993_3.pdf` (July 18, 2018).

European Commission (2020) Communication from the commission to the european parliament, the council and social committee and the committee of the regions: A european strategy for data, `https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf` (February 19, 2020).

Farronato, C. and Fradkin, A. (2018) The Welfare Effects of Peer Entry in the Accommodation Market: The Case of Airbnb, *NBER Working Paper*, **24361**.

Ferreri, M. and Sanyal, R. (2018) Platform economies and urban planning: Airbnb and regulated deregulation in london, *Urban Studies*, **55**, 3353–3368.

Fox, J. T., il Kim, K. and Yang, C. (2016) A simple nonparametric approach to estimating the distribution of random coefficients in structural models, *Journal of Econometrics*, **195**, 236–254.

Garcia-López, M.-À., Jofre-Monseny, J., Martínez-Mazza, R. and Segú, M. (2020) Do short-term rental platforms affect housing markets? Evidence from Airbnb in Barcelona, *Journal of Urban Economics*, **119**, 103278.

Geissinger, A., Laurell, C. and Sandström, C. (2020) Digital Disruption beyond Uber and Airbnb: Tracking the long tail of the sharing economy, *Technological Forecasting and Social Change*, **155**, 119323.

Grunes, A. P. and Stucke, M. E. (2015) No mistake about it: The important role of antitrust in the era of big data, *Antitrust Source (Apr. 2015), Online, University of Tennessee Legal Studies Research Paper No. 269*, available at SSRN: `https://ssrn.com/abstract=260051` (April 28, 2015).

Guda, H. and Subramanian, U. (2019) Your Uber Is Arriving: Managing On-Demand Workers Through Surge Pricing, Forecast Communication, and Worker Incentives, *Management Science*, **65**, 1995–2014.

Hagiu, A. and Wright, J. (2020) Data-enabled learning, network effects and competitive advantage, *Working Paper*.

Hall, J. D., Palsson, C. and Price, J. (2018) Is Uber a substitute or complement for public transit?, *Journal of Urban Economics*, **108**, 36–50.

Haywood, J., Mayock, P., Freitag, J., Owoo, K. A. and Fiorilla, B. (2017) Airbnb & Hotel Performance. An analysis of proprietary data in 13 global markets, Tech. rep., STR.

He, D., Kannan, A., Liu, T., McAfee, R., Qin, T. and J.M., R. (2017) Scale Effects in Web Search, In: R. Devanur N., Lu P. (eds) Web and Internet Economics. WINE 2017. Lecture Notes in Computer Science, vol 10660. Springer, Cham.

Horn, K. and Merante, M. (2017) Is home sharing driving up rents? Evidence from Airbnb in Boston, *Journal of Housing Economics*, **38**, 14–24.

Houser, D. and Wooders, J. (2006) Reputation in auctions: Theory, and evidence from eBay, *Journal of Economics & Management Strategy*, **15**, 353–369.

Hubbard, T. N. (2002) How do consumers motivate experts? Reputational incentives in an auto repair market, *The Journal of Law and Economics*, **45**, 437–468.

Huber, K., Lindenthal, V. and Waldinger, F. (2019) Discrimination, Managers, and Firm Performance: Evidence from "Aryanizations" in Nazi Germany, working Paper.

Hunold, M., Kesler, R. and Laitenberger, U. (2020) Rankings of Online Travel Agents, Channel Pricing, and Consumer Protection, *Marketing Science*, **39**, 92–116.

Jin, G. Z. and Leslie, P. (2009) Reputational incentives for restaurant hygiene, *American Economic Journal: Microeconomics*, **1**, 237–67.

Joachims, T. (2002) Optimizing Search Engines Using Clickthrough Data, in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, pp. 133–142.

Jones, C. I. and Tonetti, C. (2020) Nonrivalry and the Economics of Data, *American Economic Review*, **110 (9)**, 2819–58.

Koster, H., van Ommeren, J. and Volkhausen, N. (2018) Short-term rentals and the housing market: Quasi-experimental evidence from Airbnb in Los Angeles, *CEPR Discussion Paper*, **13094**.

Kroft, K. and Pope, D. G. (2014) Does Online Search Crowd Out Traditional Search and Improve Matching Efficiency? Evidence from Craigslist, *Journal of Labor Economics*, **32**, 259–303.

Lam, C. T., Liu, M. and Hui, X. (2020) The Geography of Ridesharing: A Case Study of New York City, *Working Paper*.

Lambrecht, A. and Tucker, C. E. (2015) Can Big Data protect a Firm from Competition?, *SSRN*, `https://ssrn.com/abstract=2705530` (December 18, 2015).

Li, L. and Xiao, E. (2014) Money talks: Rebate mechanisms in reputation system design, *Management Science*, **60**, 2054–2072.

Li, L. I., Tadelis, S. and Zhou, X. (2016) Buying reputation as a signal of quality: Evidence from an online marketplace, *NBER Working Paper Series*, **22584**.

Livingston, J. A. (2005) How valuable is a good reputation? A sample selection model of internet auctions, *Review of Economics and Statistics*, **87**, 453–465.

Luca, M. (2016) Reviews, reputation, and revenue: The case of Yelp. com, *Com (March 15, 2016). Harvard Business School NOM Unit Working Paper*.

Luca, M. and Reshef, O. (2020) The impact of prices on firm reputation, *Harvard Business School NOM Unit Working Paper*.

Luca, M. and Zervas, G. (2016) Fake it till you make it: Reputation, competition, and Yelp review fraud, *Management Science*, **62**, 3412–3427.

Lucking-Reiley, D., Bryan, D., Prasad, N. and Reeves, D. (2007) Pennies from eBay: The determinants of price in online auctions, *The journal of industrial economics*, **55**, 223–233.

Mayzlin, D., Dover, Y. and Chevalier, J. (2014) Promotional reviews: An empirical investigation of online review manipulation, *American Economic Review*, **104**, 2421–55.

McFadden, D., Tye, W. B. and Train, K. (1977) An Application of Diagnostic Tests for the Independence From Irrelevant Alternatives Property of the Multinomial Logit Model, *Transportation Research Record*, **637**, 39–46.

Melnik, M. I. and Alm, J. (2002) Does a Seller's eCommerce Reputation Matter?, *Journal of Industrial Economics*, **50**, 337–349.

Mense, A., Michelsen, C. and Kholodilin, K. A. (2017) Empirics on the causal effects of rent control in Germany, *FAU Discussion Papers in Economics*, **24/2017**.

Neeser, D. (2015) Does Airbnb Hurt Hotel Business: Evidence from the Nordic Countries, *Mimeo*.

Netflix (2020) Netflix research - Personalization & Search: Helping members discover content they'll love, `https://research.netflix.com/business-area/personalization-and-search` (last accessed: April 17, 2020).

Newman, N. (2014) Search, Antitrust, and the Economics of the Control of User Data, *Yale J. on Reg.*, **31**, 401.

Nieuwland, S. and Van Melik, R. (2020) Regulating Airbnb: how cities deal with perceived negative externalities of short-term rentals, *Current Issues in Tourism*, **23**, 811–825.

Nilsson, N. J. (2014) *Principles of artificial intelligence*, Morgan Kaufmann.

Posner, E. A. and Weyl, G. (2018) *Radical Markets, Uprooting Capitalism and Democracy For a Just Society*, Princeton University Press.

Proserpio, D. and Zervas, G. (2017) Online reputation management: Estimating the impact of management responses on consumer reviews, *Marketing Science*, **36**, 645–665.

Prüfer, J. and Schottmüller, C. (2017) Competing with Big Data, *SSRN*, `https://ssrn.com/abstract=2918726` (February 16, 2017).

Resnick, P., Zeckhauser, R., Swanson, J. and Lockwood, K. (2006) The value of reputation on eBay: A controlled experiment, *Experimental economics*, **9**, 79–101.

Schaefer, M., Sapi, G. and Lorincz, S. (2018) The effect of big data on recommendation quality: The example of internet search, *DIW Berlin Discussion Paper*, **1730**.

Schaefer, M. and Tran, K. (2020) Airbnb, hotels, and localized competition, *DIW Berlin Discussion Paper*, **1889**.

Schepp, N.-P. and Wambach, A. (2015) On Big Data and its Relevance for Market Power Assessment, *Journal of European Competition Law & Practice*, **7**, 120–124.

Seamans, R. and Zhu, F. (2014) Responses to entry in multi-sided markets: The impact of Craigslist on local newspapers, *Management Science*, **60**, 476–493.

Sokol, D. D. and Comerford, R. (2015) Antitrust and Regulating Big Data, *Geo. Mason L. Rev.*, **23**, 1129.

Stenzel, A., Wolf, C. and Schmidt, P. (2020) Pricing for the stars: Dynamic pricing in the presence of rating systems, in *Proceedings of the 21st ACM Conference on Economics and Computation*, pp. 273–274.

Sundararajan, A. (2017) *The sharing economy: The end of employment and the rise of crowd-based capitalism*, Mit Press.

Teubner, T., Hawlitschek, F. and Dann, D. (2017) Price determinants on AirBnB: How reputation pays off in the sharing economy., *Journal of Self-Governance & Management Economics*, **5**.

The Economist (2017) The world's most valuable resource is no longer oil, but data, `https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data` (May 6, 2017).

The Economist (2018) How to tame the tech titans, `https://www.economist.com/leaders/2018/01/18/how-to-tame-the-tech-titans` (January 18, 2018).

Tucker, C. (2019) Digital Data, Platforms and the Usual [Antitrust] Suspects: Network Effects, Switching Costs, Essential Facility, *Review of Industrial Organization*, **54**, 683–694.

Valentin, M. (2020) Regulating Short-Term Rental Housing: Evidence from New Orleans, *Real Estate Economics*, **forthcoming**.

Vaughan, R. and Daverio, R. (2016) *Assessing the size and presence of the collaborative economy in Europe*, Publications Office of the European Union.

Verboven, F. (1996) International price discrimination in the European car market, *The RAND Journal of Economics*, **27**, 240–268.

Yoganarasimhan, H. (2020) Search personalization using machine learning, *Management Science*, **66**, 1045–1070.

Yrigoy, I. (2019) Rent gap reloaded: Airbnb and the shift from residential to touristic rental housing in the Palma Old Quarter in Mallorca, Spain, *Urban Studies*, **56**, 2709–2726.

Zegners, D. (2019) Building an Online Reputation with Free Content, *SSRN*, https://ssrn.com/abstract=2753635 (July 27, 2019).

Zervas, G., Proserpio, D. and Byers, J. W. (2017) The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry, *Journal of Marketing Research*, **54**, 687–705.