The Distributional Impact of the Sharing Economy on the Housing Market

Sophie Calder-Wang
Harvard University *

December 27, 2019
Latest Version Here

Abstract

What is the impact of the sharing economy, pioneered by companies such as Airbnb, on the housing market? In this paper, I estimate the welfare and distributional impact of Airbnb on the residents of New York City. I develop a model of an integrated housing market, in which a landlord can offer a housing unit for rent either on the traditional long-term rental market or on the newly available short-term rental market. By estimating a structural model of residential choice and linking it to detailed Airbnb usage data, I estimate the effect of such reallocation on the equilibrium rents across different housing types and demographic groups. In addition, to evaluate the gains from direct home-sharing, I estimate a supply system featuring heterogeneous costs. Overall, renters in New York City suffer a loss of $178mm per annum, as the losses from the rent channel dominate the gains from the host channel. I find that the increased rent burden falls most heavily on high-income, educated, and white renters because they prefer housing and location amenities that are most desirable to tourists. Moreover, there is a divergence between the median and the tail, where a few enterprising low-income households obtain substantial gains from home-sharing. Thus, this paper delivers a nuanced characterization of the winners and losers of the sharing economy, and provides a framework for understanding the consequences of regulating such technological innovations.

*Department of Economics. *Email:* sophiecalderwang@g.harvard.edu. I would like to thank my advisors Ariel Pakes, Edward Glaeser, Paul Gompers, Robin Lee, and Adi Sunderam. Thanks to Alex Bell, John Campbell, Lauren Cohen, Mark Egan, Chiara Farronato, Sam Hanson, Nir Hak, Yizhou Jin, Myrto Kalouptsidi, Josh Lerner, Jing Li, Chris Malloy, David Martin, Mark Shepard, Jeremy Stein, Boris Vallee, Ron Yang for helpful conversations at various stages of the paper. This research was funded in part by the Joint Center for Housing Studies at Harvard University and the Ewing Marion Kauffman Foundation. All errors are my own.

1 Introduction

Economic theory teaches that cost-reducing new technologies should improve welfare. By substantially reducing transaction costs, platform companies such as Airbnb allow existing housing units to be used by short-term visitors in exchange for payment. Such innovation improves the allocation and utilization of the underlying asset. However, it is not necessarily Pareto improving.

As housing supply is constrained in many coastal markets in the United States,[1] there are significant concerns that Airbnb exacerbates housing affordability problems. Many worry that housing units are being reallocated away from the traditional long-term rental market and thereby displacing existing residents. Legal battles continue in places such as New York City where Mayor de Blasio signed legislation to curb Airbnb rentals in late 2018. However, the law was subsequently blocked by the court amid privacy concerns.

Proponents of Airbnb argue that the additional income that hosts earn from home-sharing, especially in expensive cities, is vital to their livelihood. An important feature of many sharing economy platforms is that services are produced by peers, rather than firms, and therefore can distribute gains directly to individuals. Therefore, the question is empirical: For New York City residents, does the welfare gain from home-sharing offset the welfare loss from increased housing costs? Moreover, how does the welfare impact differ across key demographic characteristics, such as income, education, race, and family structure?

To answer these questions, I specify and estimate a structural model that highlights two key innovations that Airbnb brings to the housing market. First, the long-term rental market and the short-term rental market become integrated on the supply side: An absentee landlord who owns a housing unit can choose between the two markets, whichever yields greater profit. In equilibrium, a fraction of the housing units are reallocated to Airbnb. Since housing supply is inelastic, such reallocation raises prices for long-term renters and decreases their welfare.[2] Second, the utilization of existing housing units increases when renters themselves offer space in their homes to host short-term visitors, especially during times when short-term rental demand is high. The proceeds from such direct home-sharing raise the welfare of these residents.

My structural model flexibly incorporates rich heterogeneity in household preferences as well as heterogeneity in the cost of home-sharing, which allows me to analyze the distributional impact across demographic groups. I model the demand for long-term rentals as a discrete choice

[1] The amount of housing construction in New York City has been depressed over the past three decades. In fact, 41% of the homes today were built prior to 1940, and 88% of the homes were built prior to 1990. Only 2.9% of the homes were built since 2010. See Figure 5 for more details.

[2] In the case of owner-occupiers, they can be thought of as renting from themselves. However, as 67% of the housing units in NYC are renter-occupied, my analysis focuses primarily on the renters and returns to owner-occupiers at the end.

problem featuring heterogeneous household preferences over housing attributes (McFadden, 1978; Bayer, Ferreira and Mcmillan, 2007). In addition to the rental price, the demand model captures a set of housing attributes ranging from hedonic attributes (such as the number of bedrooms, year built, type of structure) to neighborhood attributes (such as distance to job centers and neighborhood demographics in terms of race, ethnicity, and education). Moreover, it also allows for a horizontal preference over these attributes; For example, it captures the differential preferences over living in a predominantly ethnic neighborhood depending on the ethnicity of the household members. Compared to the traditional Alonso-Mills-Muth analysis of the urban problem (Alonso, 1964; Mills, 1967; Muth, 1969), the discrete choice framework makes explicit the preferences over a wide vector of housing attributes. Given that the entry of Airbnb varies greatly across space and housing types, the ability to incorporate heterogeneity into the long-term rental demand is crucial to evaluating its distributional impact.

I model the supply of short-term rentals by residents as a binary choice problem where a resident decides whether to share her home on a given day with a short-term visitor at the prevailing market price. She makes this decision based on the trade-off between the income she makes and the cost of providing such short-term rental services. The short-term rental supply model allows households to have differential costs based on their demographic characteristics, such as age, education, and family structure. It also allows for differential price sensitivity based on household income. To my knowledge, even though others have estimated the overall distribution of home-sharing costs (Farronato and Fradkin, 2018), this paper is the first to examine how costs of peer production differ across income and demographic characteristics.

To estimate the structural model, I adapt and extend methods from the empirical industrial organization literature. For long-term rental demand, I use the American Community Survey (ACS) Public Use Microdata Sample, whose key benefit is the availability of individual-level data. I observe the full vector of household demographic characteristics together with the full vector of housing attributes chosen by the household, including the location of the home at the neighborhood level. I construct moment conditions that match market shares, as well as the covariance between the housing attributes and the average demographic characteristics of households living in those homes (Berry, Levinsohn and Pakes, 2004). To address the concern that there may be an unobservable housing quality correlated with price, I construct an instrument based on a home's location in the characteristics space, following Bayer, Ferreira and Mcmillan (2007).

To estimate the short-term rental supply, I leverage high-frequency Airbnb transaction data. Although Berry, Levinsohn and Pakes (1995); Nevo (2000, 2001) (BLP) methods are widely used by researchers to estimate demand systems, I propose an adaptation so they can be used to estimate a heterogeneous *supply* system.[3] Since I observe the location of each housing unit on Airbnb, variations in the distribution of neighborhoods demographics allow me to estimate the heterogeneity in the cost of home-sharing. Moreover, price variations at the daily level allow me to estimate the heterogeneity in the price coefficient. In addition, to address unobservable costs, I use a measure of short-term demand seasonality as the price instrument. Since the high-frequency daily data results in a large number of market-share equations to match, I employ the MPEC procedure developed in Dubé, Fox and Su (2012) to improve the numerical performance of the estimator.

With the estimated model parameters, I conduct two counterfactual analyses to evaluate the distributional impact of Airbnb on the participants of the housing market. To evaluate the welfare impact through the rent channel in the long-term rental market, I perform a counterfactual analysis where all housing units made available by absentee landlords on Airbnb are returned to the long-term rental market. To evaluate the welfare impact via the host channel in the short-term rental market, I perform a counterfactual analysis where the hosts are no longer allowed to participate in home-sharing.[4]

The key findings are threefold. (i) The net impact of Airbnb aggregated across all renters is a loss of \$178mm per annum (p.a.), because the losses from the rent channel at \$201mm p.a. dominate the gains from the host channel at \$23mm p.a. (ii) While the median renter loses \$125 p.a., more significant losses are suffered by renters who are high-income, educated, and white, because they demand housing types that are more desirable in the short-term rental market. (iii) There is a divergence between the median and the tail: The equilibrium rent increase affects all 2.1 million renter-occupied units in New York City, but the host gains accrue heavily to a small fraction of households with particularly low costs of sharing, including low-income families.

Specifically, to compute the welfare loss through the rent channel, I link the detailed Airbnb penetration data across neighborhoods and housing types with the estimated long-term rental demand model. Combined with the assumption that the total supply of physical structures available for housing is fixed, I back out the entire vector of price changes across all housing types because of the supply squeeze due to Airbnb. Across New York City, I estimate that about 0.68% of the housing units were likely to have been reallocated,[5] resulting in an equilibrium price increase of 0.71%. The compensating variation for the median renter earning \$47,000

[3]BLP methods only require aggregate data, which is particularly useful in my setting given that part of the ongoing legal feud in New York City concerns the regulator's inability to obtain individual host-level data.

[4]Since these counterfactuals reverse the entry of Airbnb, I interpret the negative of the compensating variation therein as the welfare and distributional implications of Airbnb, as both the long-term and the short-term rental models are static. However, it ignores dynamic considerations such as switching costs. Therefore, the results here should be interpreted as a static approximation.

[5]This is based on the number of housing units marked as available on Airbnb for over 180 days in 2018.

annually is \$128 p.a. When aggregated across all renters, it amounts to a total transfer of \$200mm p.a. from renters to property owners, or \$2.7bn in NPV terms.[6] It also leads to a welfare loss of \$1mm p.a. for renters displaced from the city.

The presence of severe housing supply restrictions is the main driver for the elevated and widespread equilibrium price increase in the long-term rental market. When a number of housing units of type h leave the long-term rental market due to Airbnb, three forces act. First, because the supply of type h is inelastic, a permanent reduction leads to a price increase in h. Second, displaced renters may try to substitute to another housing type h' in the city. Since the supply of h' is also inelastic, the price of type h' also increases. Finally, the price increase in h' creates a feedback loop as some return to choose type h, thereby pushing up the price of h even further. It continues until enough renters leave the city. Such spillover effects are the primary reasons for an elevated equilibrium price response even in neighborhoods that are relatively far from the city center and have low levels of direct Airbnb activity. Conceptually, the presence of severe housing supply restrictions serves as a quantity-fixing mechanism in the housing context: When the price of a substitute product increases, the optimal response is to *increase* one's own price when the total quantity is fixed.

In terms of the distributional impact via the rent channel, I find the most significant welfare losses, when measured in dollar terms, are suffered by renters who are high-income, educated, and white.[7] I find that the median renter in the top income quintile suffers a loss of \$167 p.a., whereas the median renter in the bottom income quintile suffers \$123 p.a. In terms of education, the median renter with a college degree loses \$156 p.a., compared to a loss of \$120 p.a. for those without a college degree. Across race and ethnicity, the median white renter loses \$152 p.a., whereas the loss is \$134 p.a. for African American renters and \$113 p.a. for Hispanic renters.

Contrary to what typical political or media narratives purport, welfare differences are primarily driven by the actual geographical patterns of Airbnb usage, where its penetration in New York City tends to be higher in educated, high-income, and predominantly white neighborhoods. Although the impact on higher-income renters is exacerbated by the fact that they also tend to have a greater willingness to pay for housing amenities in general, I show that the role of geography is dominant.

Next, to compute the welfare gains via the host channel, I use the parameters estimated from the supply model to compute the compensating variation if residents no longer had the

[6] Using a capitalization rate of 7.5% based on New York City hotel REITs as of 2018, based on data compiled by CBRE Research.

[7] My focus here is to conduct a positive analysis to estimate the welfare impact in dollar terms. I do not impose any differential social welfare weights across individuals of different levels of education, race, or income. Insofar as a municipal planner wishes to use an alternative set of welfare weights, the analysis here provides the basis for which those weights could be applied.

option to share their homes. The distribution of the welfare gains has a large mass close to zero and a heavy right tail. I find that the host surpluses are irrelevant for the vast majority of the residents: The median resident gains a surplus of only \$0.4 p.a., and the 75th-percentile resident makes \$5.9 p.a., as the fraction of residents who have actually hosted guests on Airbnb remains low, averaging to only 0.8% of housing units. However, the expected gain to households in the right tail (above the 99th percentile) amounts to over \$300 p.a. When aggregated across all renters, the total host surplus produced by supplying rooms to short-term visitors on Airbnb amounts to \$23mm per year, or approximately \$300mm in NPV terms.[8]

In terms of the distributional impact via the host channel, larger gains accrue to lower-cost suppliers, which tend to be young and educated households with no children. Some low-income households also benefit more as they are better able to take advantage of peak demand in the short-term rental market. As lower-income households tend to be more price sensitive, their supply is more elastic. Conditional on being in the right tail of the host surplus distribution, households in the bottom income quintile expected a gain of \$454 p.a., compared to a gain of \$233 p.a. for the top income quintile. As a result, given the pattern of short-term rental activity, Airbnb does not necessarily exacerbate the income inequality *among* renters. However, Airbnb is likely to increase wealth inequality, depending on the status of property ownership.

Taken together, I find that the welfare loss via the rent channel at an NPV of \$2.7bn dwarfs the gains from the increased utilization via the host channel at an NPV of \$300mm. Since housing supply is inelastic, reallocation by absentee landlords raises rents on *all* units, not just for the particular unit moved to Airbnb. Nonetheless, the cost of home-sharing remains high for most people, evidenced by the fact that only a small fraction of households become hosts. As the losses from the rent channel are widespread and the gains from the host channel are concentrated, the net welfare impact remains negative for the vast majority of renters. With 67% of the housing units being renter-occupied, the median household in the city experiences a loss of \$114 p.a. As a result, even a simple voting model favors restricting Airbnb reallocation.

From the perspective of the social planner, since the rent increase results in a mere transfer to absentee landlords, the overall welfare impact of Airbnb remains positive because it also includes the economic gains accrued to all hosts, as well as the surpluses accrued to tourists net of hotel losses. However, the substantial transfer from renters to property owners reflects the regulatory conundrum caused by the severe housing supply restrictions in place. More broadly, this paper shows that taking into account the existing regulatory constraints and its ensuing market power is essential to evaluate whether gains from technological innovations could be shared equitably throughout society.

[8]Note that the surplus number estimated here is only for resident hosts sharing rooms in their homes. Absentee landlords who have reallocated to Airbnb also obtain surpluses from providing short-term rental services, presumably greater than their foregone rent. Their gains also matter to the social planner, but they are not the focus of my supply estimator.

Related Literature

The literature on the sharing economy is nascent but rapidly expanding. Theoretical work such as Filippas, Horton and Zeckhauser (2019) explore the impact of the sharing economy on asset ownership in the long-run. Within the empirical literature, many study the design of peer-to-peer platforms (Edelman, Luca and Svirsky, 2017; Jaffe, Coles, Levitt and Popov, 2019). Related to outcomes in the housing market, Horn and Merante (2017) and Barron, Kung and Proserpio (2018) provide evidence of Airbnb leading to increased rent and home prices using panel data following the entry of Airbnb. Valentin (2019) and Calder-Wang (2019) take advantage of boundary discontinuities and find evidence of depressed home prices following short-term rental restrictions in the French Quarter of New Orleans, and in Anaheim, California, respectively. Using a structural approach, Farronato and Fradkin (2018) estimate the welfare impact of Airbnb on tourists and hotels. This paper is the first to develop a structural model to quantify the distributional impact of the sharing economy on the participants of the housing market.

This paper also finds that supply constraints in the housing market (Saiz, 2010; Gyourko, 2009; Gyourko and Molloy, 2015; Glaeser and Gyourko, 2018) have a large effect on how the efficiency gains from the sharing economy are distributed, thereby complementing the existing literature on how housing constraints create economic distortions, especially in the labor market (Ganong and Shoag, 2017; Hsieh and Moretti, 2019). In addition, since the ability to rent to short-term visitors alters the cash-flow-generating ability of the underlying housing asset, this paper is also related to studies that estimate the determinants of housing amenity value, ranging from school quality, crime, nearby foreclosures, and various environmental factors.[9] Methodologically, this paper treats explicitly the fact that the housing stock is inherently heterogeneous.[10]

This paper adds to the growing literature that employs structural models to capture equilibrium effects and consumer welfare due to financial innovations. Buchak, Matvos, Piskorski and Seru (2018) show that technology provides an avenue to dampen capital requirements in mortgage lending. Higgins (2019) argues that the adoption of financial technology exhibits two-sided network effects, which benefit wealthy consumers. Koijen and Yogo (2016) estimate insurance demand and evaluate the welfare implications of regulatory arbitrage; housing reallocation to Airbnb can be viewed as a form of arbitrage brought about by technology. Rich structural models are also used to empirically assess the impact of existing or proposed fi-

[9]This is an extensive literature that employs a variety of methods, including both choice-based sorting (Ferreyra, 2007; Timmins, 2007; Banzhaf and Walsh, 2008; Bayer, Keohane and Timmins, 2009; Klaiber and Phaneuf, 2010; Tra, 2010; Galiani, Murphy and Pantano, 2015) and reduced-from estimates (Black, 1999; Greenstone and Gallagher, 2008; Campbell, Giglio and Pathak, 2011; Davis, 2011; Autor, Palmer and Pathak, 2014, 2017).

[10]The differentiated nature is also characterized in Rosen (1974); Epple (1987), where Wong (2019) provides a comparison between the hedonic marginal willingness-to-pay and those estimated by discrete choice models.

nancial market regulations, especially taking into account supply-side responses and consumer heterogeneity. For instance, Robles-Garcia (2018) and Benetton (2018) emphasize the shifting of market power among mortgage originators and brokers. Cuesta and Sepulveda (2019) study the impact of price regulation on consumer credit, whereas Nelson (2019) examines its distributional impact across different types of borrowers.

The remainder of the paper proceeds as follows. Section 1.1 starts with a stylized model, highlighting the key innovations of Airbnb. Section 2 discusses the background and the data used for the analysis. Section 3 presents the main structural model, followed by Section 4, describing how the models are estimated and the parameters obtained in . Section 5 performs the counterfactual analysis examining the welfare and distributional impact of Airbnb via the rent channel and via the host channel, respectively. Section 6 concludes.

1.1 A Stylized Model

In this subsection, I present a stylized version of the model highlighting the key innovations that Airbnb brings to the housing market. First, the long-term rental market and the short-term rental market become integrated as absentee landlords can offer their housing units in either market. Second, residents can act as peer suppliers to participate directly in the production of short-term rental services. The former improves the allocative efficiency, whereas the latter increases the utilization of existing homes. Despite its stylized nature, it illustrates the key outcomes of this paper, namely, the welfare loss through the rent channel and the welfare gain through the host channel. It also motivates my use of a model-based approach and clarifies the main assumptions made.

In Figure 1, Panel A on the top left started with the long-term rental market, whereas Panel B started with the short-term rental market. Before the arrival of Airbnb, these two markets were completely separate.[11] Both markets were in equilibrium with different market-clearing prices at p_0^L and p_0^S, respectively.[12] In other words, physical structures built for residential purposes could only be rented to long-term tenants. It used to be prohibitively costly for a property owner to rent out a residential housing unit on a short-term basis.[13]

As a large-scale home-sharing intermediary, Airbnb has rapidly brought down search costs and reduced asymmetric information between hosts and guests. An absentee landlord is no longer confined to renting in the traditional long-term rental market but gains the option to

[11]In practice, there exist residential apartments that accept guests from the corporate travel market. However, I abstract that component away, since it is much smaller than the overall housing market in NYC. I also assume that hotel operators do not accept long-term tenants.

[12]To make them comparable, the prices may be thought of as the equivalent daily rental rates in the two markets.

[13]The average length of a stay on Airbnb in New York City is 4.5 days.

participate in the newly available short-term rental market.[14] For the purpose of this stylized model alone, assume the cost of operating in either market by absentee landlords is zero. If prices are higher in the short-term rental market $p_0^S > p_0^L$, then absentee landlords will be induced to reallocate toward Airbnb and obtain higher prices. As more and more housing units are reallocated, it reduces the price wedge between the two markets. In equilibrium, a no-arbitrage condition pins down the new market price $p_1 = p_1^L = p_1^S$ that clear both markets, as well as the equilibrium number of reallocated housing units S^A.

Since housing supply has been inelastic in New York City, I assume that the total number of physical structures is fixed in the model. As such, any reduction in the number of housing units available for long-term rental results in a higher equilibrium rent $p_1^L > p_0^L$. The blue shaded regions of Panel A illustrate the welfare impact on the renters: The rectangle above the remaining renters represents the welfare transfer to property owners, whereas the triangle above the displaced renters represents the welfare loss for those who leave the city. Importantly, the detailed Airbnb usage records allow me to tabulate the reallocated quantity S^A from the data directly. As a result, once I estimate the slope of the long-term rental demand D^L, together with the observed S^A, I can compute the welfare impact of Airbnb via the rent channel.[15]

The welfare loss through the rent channel is only one part of the overall effect, since Airbnb also allows residents to increase the utilization of their homes without displacing themselves.[16] Panel D in Figure 1 illustrates the additional short-term rental supply that is provided by resident hosts, denoted as S^R. It is upward sloping because more residents will find the hassle of home-sharing worthwhile if the price p_1^S is high. Nonetheless, the no-arbitrage condition for property owners still implies that there exist an equilibrium quantity S^A and an equilibrium price $p_1 = p_1^L = p_1^S$ that clears both markets. In Panel D, the green shaded region indicates the surplus accrued to such resident hosts. Therefore, once I estimate the slope of the short-term supply S^R, I can compute their welfare gains via the host channel, which can then be netted against the welfare losses through the rent channel.

Hence, the stylized model illustrates that the key welfare outcomes (shown as the blue and green shaded regions respectively in Figure 1) can be computed by estimating the slope of the long-term rental demand D^L and the slope of the short-term rental supply S^R.

The stylized model not only demonstrates the benefits for an equilibrium model-based ap-

[14]In many cities, the legality of Airbnb has been hotly debated. In the case of NYC, although renting out a Class A unit for less than 30 days without the presence of its permanent resident is a violation of its Multiple Dwelling Law, it is also estimated that the law is not effectively enforced (Jia and Wagman, 2018).

[15]Correspondingly, in the short-term rental market, since the price has declined $p_1^S < p_0^S$, the net surplus is designated by the gray triangle, enjoyed by tourists visiting the city, and net of hotel losses.

[16]The potential resident hosts on Airbnb include both renters and owner-occupiers. I do not explicitly model whether such home-sharing violates specific leasing agreements or home-association agreements, but they are implicitly incorporated as the cost to home-sharing.

proach, it also highlights the need to incorporate heterogeneity into the full structural model. A model-based approach captures the equilibrium effects by ensuring the market clearing conditions are satisfied in both markets before and after the entry of Airbnb. Specifically, since it allows households to re-optimize when faced with a new price vector, a fully-specified long-term rental demand model can estimate the welfare change when the *bundle* of housing attributes available has changed. Moreover, a model-based approach featuring heterogeneous preferences can capture the distributional impact over relevant household characteristics, thereby evaluating the net welfare impact by household income, race, education, and family structure. This is especially relevant for those concerned about income or racial disparities. Lastly, quasi-experiments that allow researchers to identify at least the price effect of Airbnb are available only in limited settings,[17] leaving the welfare and distributional issues in some of the most important housing markets unanswered.

The stylized model also clearly shows the limitations and assumptions used. First, when computing the equilibrium price impact of Airbnb via the rent channel, the total supply of physical structures is assumed to be fixed. Another assumption is the absence of negative externalities of short-term visitors on the neighborhood, potentially due to increased noise and traffic. Also, an implicit assumption here is that the ability of residents to act as peer suppliers do not alter their long-term rental demand, which is an empirical simplification.[18] Finally, the tourist welfare will not be explicitly modeled, as the paper focuses on the residents.

2 Background and Data

The combination of mobile technology and the improving designs of the reputation systems on two-sided platforms have greatly facilitated the development of the "sharing economy." Although there is no single official definition over what the sharing economy is, several important features stand out: First, it allows existing asset owners to increase utilization by allowing someone else to use their asset temporarily in exchange for payment. The growth and the development of a large-scale intermediary drastically lowers search costs. Second, the platform companies that facilitate the exchange typically do not own the assets themselves, so the services offered on the platform are fulfilled by peers. Although there exists an extensive literature on estimating the production function for firms, it is not immediately clear how it extends to such decentralized peer-production processes. In addition, the development of a well-functioning reputation system alleviates the problem of asymmetric information. Even though the buyer and the seller are typically only engaged in a one-time exchange, they interact with the platform

[17]Without natural experiments such as Valentin (2019) or Calder-Wang (2019), correlating home price changes with the growth of Airbnb's is susceptible to the usual endogeneity problem, namely that improvements in unobserved neighborhood quality drive both the short-term rental demand and the long-term rental demand.

[18]Given that the expected gains from hosting turn out to be immaterial for the vast majority of households, this assumption is a simplification. I also discuss its likely impact at the end of the paper. In other places where the expected host gains are large, such as vacation markets, this assumption may be more important.

repeatedly, thus allowing the platform to aggregate relevant user information over time.

One of the most prominent businesses in this category is the home-sharing platform Airbnb, which is an online marketplace for arranging short-term lodging, especially homestays. According to its website, the founders of Airbnb started the company in 2008 by allowing guests to sleep on their air mattress in San Francisco after they noticed that the participants of a conference were struggling to find accommodation as local hotels were sold out.[19] The novelty and the scalability of the business model has attracted over $4.4bn in venture capital funding. As of 2019, Airbnb boasts over 7 million listings across the globe, which is significantly higher than any hotel group.[20]

Figure 2 shows the rapid growth of Airbnb across cities in the United States. Between 2014 and 2018, the number of reservations made on Airbnb quadrupled in many cities. The largest metropolitan market in the U.S. measured by days reserved is New York, followed by Los Angeles and Miami. Table 1 shows that Airbnb booked 5.8 million days of stay in New York City in 2018, which is about 15% of the total number of hotel stays. The average price of an entire home on Airbnb is $224 per night, and the average price of a private room is $86 per night. In 2018, 74,963 listings had active transactions on Airbnb , representing about 2.2% of all housing units. Moreover, Airbnb is by far the most dominant player among short-term rental platforms, capturing over 90% of the market share in New York City.[21]

The average level of Airbnb activity in the city masks the extensive geographic heterogeneity across neighborhoods and boroughs, as illustrated in Figure 3. In Brooklyn, the neighborhood with the highest proportion of housing units active on Airbnb in 2018 is Greenpoint & Williamsburg (9.3%), followed by Bushwick (8.5%). In Manhattan, Chelsea, Clinton & Midtown (6.9%) is at the top, followed by Chinatown & Lower East Side (6.5%). In Queens, the most active neighborhood is Astoria & Long Island City (2.8%). In the Bronx, the most active neighborhood is Concourse, Highbridge & Mount Eden (0.4%), which has much less penetration than the other boroughs.[22]

Just because a housing unit has experienced a booking on Airbnb, it is not obvious what its alternative use would have been if Airbnb had not been invented. Importantly, there are vast variations in terms of a property's calendar availabilities. Figure A.17 shows the distribution of properties by the fraction of its calendar marked available. The distribution is spread out between 0 and 1, but a large mass of 35% are marked as available on Airbnb for over 90% of the calendar days, which indicates they are likely dedicated for short-term rental use only.

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

[20] The largest hotel company in the world, Marriott International operates approximately 1.1 million rooms.

[21] The second largest player in New York City is HomeAway. Given Airbnb's dominance in NYC, I refer to Airbnb and home-sharing platforms synonymously through the paper.

[22] I have excluded Staten Island, as it is much less likely to be in the choice set of a short-term traveler visiting NYC.

Based on calendar availability and the listing type, I approximate the proportion of housing units likely to have been reallocated away from the long-term rental market.[23] Specifically, I designate entire homes available on Airbnb for over 180 days in 2018 as being reallocated by their property owners, whose alternative use would have been long-term rentals. It averages to 0.68% of the rental housing stock, while the distribution also exhibits remarkable heterogeneity across neighborhoods. Figure 4 shows the geographic variations of such reallocation follow a similar pattern as the overall level of Airbnb activity in Figure 3, but with an even higher concentration in Manhattan. The neighborhood that experiences the highest levels of Airbnb reallocation is Chelsea, Clinton & Midtown (3.4%), followed by Murray Hill, Gramercy & Stuyvesant Town (2.4%), and then by Battery Park City, Greenwich Village & Soho (2.3%). In Brooklyn, Williamsburg & Greenpoint (1.9%) and Bedford-Stuyvesant (1.8%) are affected the most.

Data

The primary data source is a full sample of Airbnb listings scraped by a third-party data vendor, AirDNA. AirDNA started scraping the entire website of Airbnb.com comprehensively in late 2014, which is when my data start.[24] For each listing in New York City, the dataset contains detailed information about the property characteristics, including the type of property and the hedonic attributes, such as the number of bedrooms, the number of bathrooms, and other relevant amenities. Broken down by listing type, 51% are entire homes, 45% are private rooms, and less than 5% are shared rooms, as summarized in Table 1. Importantly, the dataset also scrapes the latitude and the longitude of the property,[25] which allows me to map to its corresponding neighborhood.

Another beneficial feature of the dataset is its high-frequency panel, available at the most detailed daily level. It allows me to capture the seasonal nature of the short-term rental market. For each listing and every day, I observe whether the listing was available on Airbnb, its listed price, as well as whether a reservation occurred on that day. Figure A.18 and Figure A.19 show the time series plot of the total daily quantity and the average transaction price of all private rooms in New York City. There are strong patterns of seasonality. The average transaction

[23] These are the housing units most likely to be the target of the legislative efforts signed by Mayor de Blasio on August 6, 2018. Int. 981-A requires online short-term rental platforms to report data about transactions. According to Council Member Carlina Rivera, such reported data could be used by the Mayor's Office of Special Enforcement to "pursue more effective oversight and action over the bad actors that exist throughout this largely unmonitored market." https://www1.nyc.gov/office-of-the-mayor/news/398-18/

[24] Because hosts need to make their listing information publicly available online to all potential guests, the scraper downloads all the available information. Since the scraper visits each listing multiple times per week, the number of days reserved is then backed out from changes in the calendar availabilities.

[25] Even though the scraper downloads the exact location as shown on the Airbnb website, Airbnb is known to add some noise to ensure host privacy. However, based on Wachsmuth and Weisler (2018), the perturbation is between 0 and 500 feet, which introduces minimal noise in terms of assigning a property to its neighborhood at the Public Use Microdata Area (PUMA) level.

price in the sample period for a private room is \$73, with an annualized volatility of 42%. The peak demand predictably happens on New Year's Eve each year, with its price averaged at \$94. On the other hand, the trough season happens predictably in January and February with depressed quantities and prices. The peak-to-trough ratio of the daily quantity is 3.5x, whereas the peak-to-trough ratio of the average daily price is 1.6x. Overall, the high-frequency dataset allows me to take advantage of the seasonal variations in the short-term rental demand from visitors.

The second dataset is the American Community Survey (ACS) Public Use Microdata Sample. As it contains individual-level data, it is particularly helpful in estimating the housing choice model. I observe the full vector of household demographic characteristics, together with the full vector of housing attributes chosen by the household. The key demographic variables include household income, education, race, ethnicity, age, and family size. The key housing attributes include monthly price,[26] number of bedrooms, building age, and type of building. Moreover, I also observe the location of the home at the neighborhood level.[27] For each neighborhood, I also obtain average neighborhood characteristics, including race, education, and average commute time.

Overall, New York City has 3.14 million occupied housing units, of which renters occupy over 67%. Among the renters, there are substantial demographic disparities across neighborhoods and boroughs. Table 2 shows the median renter in Manhattan makes \$67,000 a year, while the median renter in Queens makes \$52,000, and in Brooklyn \$44,000. In contrast, the median renter in the Bronx makes only \$31,000. Similar patterns of disparity exist for education. In terms of race and ethnicity, the Bronx and Brooklyn have much higher proportions (greater than 35%) of African American households than Manhattan and Queens. Meanwhile, the Bronx and Queens have higher proportions of Hispanic households than Manhattan and Brooklyn.

The geographic variations of the household demographics allow me to evaluate the distributional impact of Airbnb across these demographic characteristics. The reduced-form correlations in Table 3 reveal that there are more Airbnb listings in neighborhoods with more white, educated, and higher income residents. However, a full model is needed to translate this empirical pattern into welfare, which takes into account the equilibrium effects before arriving at the distributional implications.

[26] For renters, I observe the monthly rent. For owner-occupiers, I impute an equivalent monthly user-cost based on reported home values, as done by Bayer, Ferreira and Mcmillan (2007).

[27] In the ACS microdata, the location of the home is available at the PUMA level. Fortunately, densely populated New York City contains 55 PUMAs and they could be used as an approximation for neighborhoods. I use the 2017 1-Year Estimates to estimate the long-term rental demand model, which is a 1% sample based on the 2010 U.S. Census. I use other years from 2010 to 2017 for variables involving neighborhood changes. The estimates remain robust if I use the 5-Year Estimates.

Lastly, I augment the analysis with data from STR, which tracks supply and demand for hotels. Specifically, I obtain the daily aggregate number of hotel rooms sold in New York City from 2007 to 2018. It is used to construct a seasonality-based demand shifter to estimate the cost of peer supply.

3 Model Description

3.1 A Model of an Integrated Housing Market

In this section, I describe the main components of the model. I specify demand and supply in both the long-term and the short-term rental market, highlighting the key innovations brought about by Airbnb. First, the long-term and the short-term rental market become integrated on the supply side, improving allocative efficiency. Second, the utilization of existing housing units is increased when residents themselves offer space in their homes to host short-term visitors, especially during times when short-term rental demand is high.

A. Long Term Rental Demand

I start with a model of residential housing choice using a random utility framework following McFadden (1978) and Bayer, Ferreira and Mcmillan (2007).[28] Each resident faces a discrete choice problem among all housing types in the city, and each housing type is defined by the neighborhood n in which the housing unit is located, as well as the physical characteristics of the unit, which include the number of bedrooms, age of the building, and indicators for the type of the building. As a result, the housing stock in New York City is divided into 1,050 such types. Within each housing type, there are N_h units that are not further differentiated beyond an idiosyncratic component $\epsilon^L_{i,j}$. Hence, the long-term rental utility of household i derived from housing unit j of type h is

$$u^L_{i,j} = \alpha^L_i p^L_h + \boldsymbol{\beta}^L_i \mathbf{X}^L_h + \xi^L_h + \epsilon^L_{i,j} \tag{3.1}$$

where the superscript L indicates quantities pertaining to the long-term rental market. p^L_h refers to the price of housing type h, and $\mathbf{X}^L_h$ includes both the physical and the neighborhood characteristics. The neighborhood characteristics include the percentage of the households with a college degree and the percentage of African American, Hispanic, and Asian households. I also include a measure of its location amenity using the average commuting time.[29] In addition,

[28] I use the phrase "long-term rental" to capture the residential housing choices made by long-term residents. It is to be contrasted with the "short-term rental" market where visitors have a demand for housing for typically much shorter periods (e.g. at the daily level). Even though the exposition of the model appears to assume that all households are renters, it can be readily extended to incorporate owner-occupiers in the sense that the model captures the equivalent user-cost of the owner-occupiers.

[29] In the baseline model, I use the average commuting time of workers living in the neighborhood. In a richer model, since I observe the work location at the borough level, the location amenity could be modeled more finely as household-specific.

I allow an unobserved quality component ξ_h^L that could be correlated with price. The price coefficient α_i^L and the coefficients on the housing characteristics β_i^L are determined in a flexible way based on the vector of observable demographics z_i, including household income, race, ethnicity, education, and family size.[30]

Each household i makes an optimal housing choice j by maximizing utility:

$$y_i^L = j \iff u_{i,j}^L > u_{i,-j}^L$$

The set of households who choose housing type h is simply the union of those who choose any housing unit j that is of type h:

$$A_h^L = \bigcup_{j:h(j)=h} \{z_i, \epsilon_{i,\cdot}^L : u_{i,j}^L > u_{i,-j}^L\}$$

The total long-term rental demand for housing type h can be obtained by integrating over all households

$$D_h^L(p_h^L, p_{-h}^L) = \int_{A_h^L} dP(\epsilon^L) dP_D^*(z)$$

where P_D^* represents the empirical distribution of demographic characteristics of all potential city residents.[31]

B. Long-Term Rental Supply

The key assumption here is that the total number of housing units of each type in the city is fixed at S_h^F. In a market without the home-sharing platform, the city's housing market would be fully characterized by the market clearing price for each housing type h in the long-term rental market:

$$\forall h : \quad D_h^L(p_h^L, p_{-h}^L) = S_h^F$$

However, with the home-sharing platform, I will proceed to specify the short-term rental supply before imposing market clearing.

C. Short-Term Rental Supply

I categorize the supply of short-term rentals on Airbnb into two different types: those provided by absentee landlords, and those provided by residents.

[30] The model assumes that the unobserved quality is vertical. This assumption is more reasonable when the vector of observable X_h and the vector of demographics z_i are sufficiently rich to capture horizontal sorting.

[31] The market size for living in a city is defined as the relevant metro market, which includes the city itself and the surrounding areas within a commutable distance. In practice, I focus on all the contiguous counties that surround New York City, namely Hudson, Nassau, Westchester, and Bergen.

The first type is provided by absentee landlords who reallocated housing units from the long-term to the short-term rental market. In this case, for every unit reallocated, there is one fewer unit available for long-term tenants.

The second type is provided by the residents directly. In this case, the resident of the housing unit chooses to supply space to the short-term rental market if she finds that the short-term rental income in a given period is greater than the value for her alternative personal use. This second type of short-term supply increases the utilization of housing units already occupied by residents.[32]

C1. Short-Term Rental Supply from Absentee Landlord

With Airbnb, an absentee landlord owning a housing unit may now consider short-term rental as an alternative to long-term rental. For each housing type h on day t, the price of the unit in the short-term rental market is $p^A_{h,t}$.

The utility of accepting a short-term rental visitor on day t in housing unit j of type h is:

$$u^A_{j,t} = p^A_{h,t} + \nu^A_{j,t} \tag{3.2}$$

where $\nu^A_{j,t}$ represents the cost of operating the short-term rental (including cleaning and providing supplies, for instance). However, the decision to reallocate away from the long-term rental is made with the consideration of a much longer time period T (e.g. a year). Hence, an absentee landlord chooses to reallocate if he can make more money in the short-term rental market overall:

$$y^A_j = 1 \iff \frac{1}{T}\sum_{t=0}^{T} \max\{\, p^A_{h,t} + \nu^A_{j,t},\, 0\,\} > p^L_h + \nu^L_j$$

where ν^L_j is the cost associated with operating property j in the long-term rental market and p^L_h is the long-term rental rate per day.[33] Hence, the total number of housing units of type h that is reallocated by absentee landlords is as follows

$$S^A_h(p^L_h, p^A_{h,\cdot}) = \int_{A^A_h} dP(\nu^A, \nu^L), \quad A^A_h = \bigcup_{j:h(j)=j} \{\nu^A_{j,\cdot}, \nu^L_j : \quad y^A_j = 1\}$$

[32] For example, residents may choose to rent out their guest bedroom when they do not have friends or family visiting. In other words, the room would not otherwise be available as a long-term rental unit for another household. Hence, in the baseline model, the supply of short-term rental space offered by residents does not crowd out available space for long-term rental. In the long run, households may adjust their long-term rental demand due to their expected short-term rental income. I consider its likely impact in the last section of the paper.

[33] This specification abstracts away from discount rates, uncertainty, and any dynamic considerations such as the option value to convert in the future. In other words, the baseline model is static, but it still allows for short-term rental vacancies due to predictable price variations from short-term rental seasonality.

On a given day t, the total short-term supply is simply the cumulative distribution of those who can operate profitably at the market rate, given that it has already been reallocated to Airbnb:

$$S^A_{h,t}(p^L_h, p^A_{h,\cdot}) = \int_{A^A_{h,t}} dP(\nu^A, \nu^L), \quad A^A_{h,t} = \bigcup_{j:h(j)=j} \{\nu^A_{j,\cdot}, \nu^L_j : \quad y^A_j = 1, \quad u^A_{j,t} > 0\}$$

C2. Short-Term Rental Supply from Residents

As the second type of short-term rental supply, city residents can directly host short-term visitors in their homes, which is a more accurate reflection of the spirit of the "sharing economy." In this case, the utility derived from the supply of a short-term rental room by household i in neighborhood n on day t is

$$u^R_{i,t} = \alpha^R_i p^A_{n,t} + \boldsymbol{\beta}^R_i \mathbf{X}^R_{n,t} + \xi^R_{n,t} + \epsilon^R_{i,t} \tag{3.3}$$

where $p^A_{n,t}$ denotes the price of an Airbnb private room and $\mathbf{X}^R_{n,t}$ denotes a number of observable shifters, importantly including a constant term representing the negative cost of providing a room. The coefficient β^R_i in front of the constant is modeled as a function of household demographics, including age, education, income, and family structure. The price coefficient α^R_i is modeled as a function of household income. The model allows for an unobserved cost at the neighborhood-day level that may be correlated with the prevailing market price $p^A_{n,t}$. A household-specific idiosyncratic taste for home-sharing on a given day is included as $\epsilon^R_{i,t}$.

As an additional housing type, the product here is a private room for short-term use. Therefore, it is only differentiated at the neighborhood and day level.[34] Thus, a given household chooses to supply a room on a given day if the utility of hosting is greater than the outside option, which is normalized to zero. The total short-term rental supply by residents in neighborhood n on day t is the integral of such households in the neighborhood:

$$S^R_{n,t}(p^A_{n,t}) = \int_{A^R_{n,t}} dP(\epsilon^R) dP^*_{D_n}(z), \quad A^R_{n,t} = \left\{z_i, \epsilon^R_{i,t} : u^R_{i,t} > 0\right\}$$

where $P^*_{D_n}(\cdot)$ denotes the empirical distribution of the household demographics in neighborhood n.[35]

[34] In other words, a private short-term rental room in a two-bedroom home is not differentiated from that in a three-bedroom home. Moreover, the estimation of the host surplus focuses on the gains derived from such private rooms as opposed to entire homes, as the private rooms are more likely to remain legal under the current regulation in New York City, which requires the permanent resident to be present.

[35] The distribution of the neighborhood demographics D_n is considered constant from day to day and thus does not have a subscript t.

D. Short-Term Rental Demand

The demand for short-term rental is characterized by a discrete choice problem where the choice set includes all short-term rental housing types available on the home sharing platform (including both entire homes of all types h and private rooms in all neighborhoods n) on a given day t, denoted by $D^A_{h,t}(p^A_{h,t}, p^A_{-h,t})$.[36]

E. Market Equilibrium

The long-term rental market is characterized by a sorting equilibrium. Namely, the equilibrium price vector p^L_h for all housing types ensures that the demand for each housing type h equals its supply, which is the number of the underlying housing units less what have been reallocated to the short-term rental market:

$$\forall h: \quad D^L_h(p^L_h, p^L_{-h}) = S^F_h - S^A_h(p^L_h, p^A_{h,\cdot}) \tag{3.4}$$

The short-term rental market is analogously characterized by the market-clearing of each short-term rental lodging option every day. Namely, the equilibrium price vector $p^A_{h,t}$ for all short-term lodging options h ensures that the short-term rental demand equals the short-term rental supply. For notational simplicity, the short-term lodging options h include all the housing types in the long-term rental market, as well as a private room in neighborhood n.

$$\forall h, t: \quad D^A_{h,t}(p^A_{h,t}, p^A_{-h,t}) = S^A_{h,t}(p^L_h, p^A_{h,\cdot}) + S^R_{h,t}(p^A_{h,t}) \tag{3.5}$$

Bayer and Timmins (2005) provide the regularity conditions that guarantee the existence of the sorting equilibrium, namely that the errors ϵ^L and ϵ^A are drawn from a continuous and well-defined distribution function.[37]

4 Estimation and Results

In this section, I provide details on how the structural models are estimated. First, I estimate the long-term rental demand model using individual-level data from the cross-section of housing choices. Then, I construct my estimator for the short-term rental supply model, using aggregate data across multiple neighborhoods and time periods. In both cases, I highlight

[36]Thus, $D^A_{h,t}$ represents the residual demand for Airbnb after hotels. Because the focus of the paper is not on the welfare of short-term visitors, I do not explicitly estimate their utility. However, one can think of a visitor's utility for a particular short-term rental option being determined by its price and its product characteristics. Then, she makes the optimal choice among them.

[37]In general, the uniqueness of the sorting equilibrium in the long-term rental market is not guaranteed if a household's long-term rental utility is affected by the choice of other individuals (e.g. preference for neighborhood racial and ethnic composition). Nonetheless, Bayer and Timmins (2005) also show that uniqueness becomes easier to sustain when there is a large number of available choices, which is likely to hold in the residential choice setting.

the moment restrictions needed and the identifying assumptions used. Lastly, I discuss the parameters estimated and the corresponding elasticities.

4.1 Estimating the Long-Term Rental Demand

In this section, I describe the two-step procedure for estimating the long-term rental demand using two sets of moment conditions.

In the first step, I estimate the heterogeneous coefficients by taking advantage of the individual-level choice data. I construct moment conditions that match market shares, as well as the covariance between the housing attributes and the demographic characteristics of the households living in those homes. In the second step, I estimate the linear coefficients. To address the concern that there may be an unobservable housing quality correlated with price, I construct a price instrument based on a home's location in the characteristics space while setting the unobserved quality to zero, following Berry, Levinsohn and Pakes (1999) and Bayer and Timmins (2007).

Recall that the utility for a household i considering home j of type h depends on its price p_h^L and its vector of housing attributes X_h^L.[38] Since α_i^L and $\boldsymbol{\beta}_i^L$ are specific to each household i, I parameterize them as the sum of two parts: The first part is common to all households, and the second part is a function of its observable demographic characteristics z_i, where the matrix π^L is fully saturated.

$$u_{i,j}^L = \alpha_i^L p_h^L + (\boldsymbol{\beta}_i^L)^T \mathbf{X}_h^L + \xi_h^L + \epsilon_{i,j}^L$$

$$\begin{bmatrix} \alpha_i^L \\ \boldsymbol{\beta}_i^L \end{bmatrix} = \underbrace{\begin{bmatrix} \alpha^L \\ \boldsymbol{\beta}^L \end{bmatrix}}_{\substack{\text{common to all;} \\ \text{part of the mean utility } \delta_h^L}} + \underbrace{\begin{bmatrix} \pi_{\alpha,1}^L \dots \pi_{\alpha,K}^L \\ \pi_{\boldsymbol{\beta},1}^L \dots \pi_{\boldsymbol{\beta},K}^L \end{bmatrix}}_{\substack{\text{household-specific;} \\ \text{part of heterogeneous utility } \lambda_{i,h}^L}} \begin{bmatrix} z_{i,1} \\ \vdots \\ z_{i,K} \end{bmatrix}$$

Denote $N_h = S_h^F - S_h^A$ as the number of housing units of type h.[39] Assuming the error ϵ^L is i.i.d. type I extreme value, I compute the choice probability analytically:

$$Pr(y_i^L \in h;\, \delta^L, \pi^L) = \frac{N_h \exp\left(\delta_h^L + \lambda_{i,h}^L\right)}{\sum_{h'} N_{h'} \exp\left(\delta_{h'}^L + \lambda_{i,h}^L\right)}$$

where δ_h^L is the mean utility from housing type h that is common to all households and $\lambda_{i,h}^L$ is

[38] I index the housing attributes by b, but omit it for simplicity when it does not cause confusion.

[39] Note that one of the housing types is the outside option, whose utility is normalized to zero.

the heterogeneous part of the utility specific to household i:

$$\lambda_{i,h}^L = \Big(\sum_k \pi_{\alpha,k}^L z_{i,k}\Big) p_h^L + \Big(\sum_k \pi_{\boldsymbol{\beta},k}^L z_{i,k}\Big)^T \mathbf{X}_h^L$$
$$\delta_h^L = \alpha^L p_h^L + (\boldsymbol{\beta}^L)^T \mathbf{X}_h^L + \xi_h^L$$

Notice that the choice probability is well-defined as long as the parameters δ^L and π^L are provided.

Step 1 Moments

In the first step, I construct moments that identify the heterogeneous parameters π^L and the mean utility δ^L.

Since I observe the actual individual-level choices $\mathbb{1}\{y_i^L \in h\}$ from the data, I construct moment conditions that match the market shares, as well as moment conditions that match the covariance between the housing attribute $X_{b,h}^L$ and the average characteristics z_k of households who choose h. For instance, I match the covariance between the number of bedrooms and the average size of the families who choose that type of house. Thus, the set of moment conditions that pin down δ^L and π^L are as follows:

$$\forall h: \quad \mathbb{E}\big[\, Pr(y_i^L \in h;\, \delta^L, \pi^L)\,\big] \quad = \quad s_h^L \tag{4.1}$$
$$\forall b,k: \quad \mathrm{Cov}\Big(\, \mathbb{E}[z_k | y_i^L \in h;\, \delta^L, \pi^L],\ X_{b,h}^L\,\Big) \quad = \quad \mathrm{Cov}\Big(\, \bar{z}_{k,h},\ X_{b,h}^L\,\Big) \tag{4.2}$$

where the left-hand side denotes the model predictions and the right-hand side is estimated from its empirical counterparts:

$$\hat{s}_h^L = \frac{1}{N}\sum_i \mathbb{1}\{y_i^L \in h\}, \quad \hat{\bar{z}}_{k,h} = \frac{1}{N}\sum_i \mathbb{1}\{y_i^L \in h\} z_{i,k}$$

where $N = \sum_{h'} N_{h'}$. Also, note that the market share moments Eq (4.1) take the expectation over all households i. The attribute covariance moments Eq (4.2) take the expectation over all housing types h, weighted by the probability that a housing type h is chosen.[40]

The identification here relies only on individual rationality, together with the fact that the housing prices and the attributes p_h^L, X_h^L are exogenous from the perspective of a single household making the choice. Since each household is assumed to be infinitesimal, this condition holds.

[40] The formulation here more closely follows the intuition provided in Berry, Levinsohn and Pakes (2004). Alternatively, Bayer, Ferreira and Mcmillan (2007) formulate an analogous set of moments using a maximum likelihood estimator, setting its score to zero. In particular, the market share moments correspond to the derivative of the log-likelihood function with respect to the mean utility δ_h^L. The attribute covariance moments correspond to the derivative with respect to each heterogeneous parameter $\pi_{b,k}^L$.

Step 2 Moments

In the second step, I construct moments that identify the linear parameters α^L and β^L as well as the unobserved quality ξ_h^L. Because I allow the market price p_h^L to be correlated with the unobserved quality ξ_h^L, I employ an identification strategy that takes advantage of the shape of the housing characteristics space, following Berry, Levinsohn and Pakes (1999) and Bayer, Ferreira and Mcmillan (2007).

Intuitively, a home that is situated in a crowded part of the housing attribute space has a low equilibrium price, regardless of its own unobserved quality. As such, I construct a price instrument using the characteristics of other homes in the city. Specifically, to characterize the impact of the attribute space on market prices, I compute an alternative vector of equilibrium prices p^{IV} as an instrument for the observed prices p^L by setting the unobserved quality to zero $\xi_h^L = 0$ and resolving the market-clearing conditions across all home types. Thus, the moment conditions for estimating the linear parameters are as follows:[41]

$$\forall h: \quad \mathbb{E}\big[Pr\big(y \in h;\, (p^{IV}, \alpha^L, \beta^L, \xi^L = 0;\, \pi^L, X^{\text{exog}})\big)\big] = s_h^L \tag{4.3}$$

$$\forall h: \quad \mathbb{E}\big[Pr\big(y \in h;\, (p^L, \alpha^L, \beta^L, \xi^L;\, \pi^L, X)\big)\big] = s_h^L \tag{4.4}$$

$$\mathbb{E}\big[\xi^L p^{IV}\big] = 0 \tag{4.5}$$

$$\mathbb{E}\big[\xi^L X_b^{\text{exog}}\big] = 0 \tag{4.6}$$

Importantly, the construction of the alternative equilibrium price p^{IV} as an instrument requires a supply-side pricing equation. Unlike a typical Nash pricing equation, the supply-side pricing equation in the housing context is a market clearing condition with a fixed housing supply, as stated in Eq (4.3).[42] Notice that for each guess of the demand parameter α^L, β^L, ξ^L, there exists a corresponding p^{IV}. Hence, the demand parameters and the price instrument are estimated jointly[43] so that all moment conditions listed above are satisfied.

The identification assumption needed is that there exists a subset of housing characteristics $\mathbf{X}^{\text{exog}}$ that is independent of the unobserved quality:

$$\xi_h^L \perp\!\!\!\perp \mathbf{X}^{\text{exog}}$$

Since p^{IV} is constructed by setting the unobserved component ξ^L to zero, then by construction, they are uncorrelated. In the housing context, the challenge is to find a subset of the housing

[41]Here, I use the heterogeneous parameters π^L that are estimated from Step 1 in the choice probability.

[42]There are other ways to form instruments based on the shape of the characteristics space, such as those used in Berry, Levinsohn and Pakes (1995). However, using a single price instrument p^{IV} to capture the equilibrium price impact of the attribute space is an approximation of the optimal instrument in the sense of Chamberlain (1987), Reynaert and Verboven (2014), and Berry, Levinsohn and Pakes (1999).

[43]In practice, an iterative procedure is used to find the fixed point.

attributes that may be considered reasonably independent of the unobserved ξ^L. Following Bayer, Ferreira and Mcmillan (2007), I use a vector of immutable attributes of the housing stock.[44] For instance, one may not think a two-bedroom home is necessarily of higher or lower unobserved quality when compared to a one-bedroom home. However, I exclude average neighborhood demographics, since it is likely that education and race affect unobserved housing quality.[45] In other words, the price instrument is computed by resolving the pricing equation Eq (4.3) with only the exogenous housing attributes.

Step 1 and Step 2 are performed sequentially to estimate the vector of heterogeneous coefficients and the linear coefficients. When the demographics vector z_i is normalized to have mean zero, the ratio of the linear coefficient over the price coefficient $-\beta_b^L/\alpha^L$ represents the willingness-to-pay of the average household for attribute X_b^L.

4.2 Estimating the Short-Term Rental Supply

In this section, I describe how I estimate the short-term rental supply. Although BLP methods are widely used by researchers to estimate demand systems (Berry, Levinsohn and Pakes, 1995; Nevo, 2000, 2001), I propose an adaptation so that it can be used to estimate a random-coefficient *supply* system. My adaptation takes advantage of the fact that the location of a housing unit on Airbnb is observed, allowing me to match the "market shares" of home-sharing supply in each neighborhood every day. As such, variations in the distribution of demographic characteristics across neighborhoods and variations due to short-term demand seasonality allow me to estimate the heterogeneity in cost. Lastly, since the high-frequency daily data results in a large number of market-share equations to match, I employ the MPEC procedure developed in Dubé, Fox and Su (2012) to improve the numerical performance of the estimator.

Recall that the utility that a resident host i living in neighborhood n derives from sharing a private room on Airbnb on day t depends on how she values the income from sharing $p_{n,t}^A$ compared to how costly it is to provide such short-term rental services. In order to capture the heterogeneity, I allow α_i^R and $\boldsymbol{\beta}_i^R$ to be household-specific. I parameterize them as the sum of

[44] The attributes included are indicators of the number of bedrooms, type of building, age of building, average commuting time to work centers, and an indicator for being inside the city.

[45] Relatedly, since I do not acquire additional instruments for these endogenous neighborhood attributes, the model only estimates the linear coefficients β^L in front of the exogenous ones. In practice, one may think of the estimated $\hat{\xi}^h$ as the sum of the true unobserved ξ^h and the endogenous component $\beta^L X^{\text{Endo}}$, which I assume to be unchanged in the subsequent counterfactuals.

a common component and a component that depends on its demographic characteristics.

$$
\begin{aligned}
u_{i,t}^R &= \alpha_i^R p_{n,t}^A + (\boldsymbol{\beta}_i^R)^T \mathbf{X}_{n,t}^R + \xi_{n,t}^R + \epsilon_{i,t}^R \\
\begin{bmatrix} \alpha_i^R \\ \boldsymbol{\beta}_i^R \end{bmatrix} &= \underbrace{\begin{bmatrix} \alpha^R \\ \boldsymbol{\beta}^R \end{bmatrix}}_{\text{common to all}} + \underbrace{\begin{bmatrix} \pi_{\alpha,1}^R \dots \pi_{\alpha,K}^R \\ \pi_{\boldsymbol{\beta},1}^R \dots \pi_{\boldsymbol{\beta},K}^R \end{bmatrix}}_{\text{household-specific}} \begin{bmatrix} z_{i,1} \\ \vdots \\ z_{i,K} \end{bmatrix}
\end{aligned}
$$

where $p_{n,t}^A$ denotes the prevailing price of an Airbnb private room in neighborhood n on day t. $\mathbf{X}_{n,t}^R = [\,1,\ t,\ t^2, Z_{month}, Z_{dow}, Z_{holiday}]^T$ captures features that contribute to the cost of sharing the room. It includes a constant term, a quadratic time trend, and dummies for month fixed effects, day-of-the-week fixed effects, and holiday fixed effects. $\xi_{n,t}^R$ represents an unobserved cost that varies at the neighborhood-day level.[46]

Since the constant term captures the (negative) cost of sharing a room such as the time and the hassle, I parametrize the coefficient β_i^R in front of it as a linear function of household income, age, education, and family structure. Also, I allow the price coefficient α_i^R to be a function of household income, permitting one's price sensitivity to differ by income.[47]

Because each resident is faced with a binary choice between sharing and not sharing, assuming the error is distributed as type I extreme value, I can analytically derive the quantity supplied:

$$
s_{i,n,t}^R(\delta_{n,t}^R, \pi^R) = \frac{\exp(\delta_{n,t}^R + \lambda_{i,n,t}^R)}{1 + \exp(\delta_{n,t}^R + \lambda_{i,n,t}^R)}
$$

where the mean utility and the heterogeneous utility are defined as follows

$$
\begin{aligned}
\lambda_{i,n,t}^R &= \Big(\sum_k \pi_{\alpha,k}^R z_{i,k}\Big) p_{n,t}^A + \Big(\sum_k \pi_{\boldsymbol{\beta},k}^R z_{i,k}\Big)^T \mathbf{X}_{n,t}^R \\
\delta_{n,t}^R &= \alpha^R p_{n,t}^A + (\boldsymbol{\beta}^R)^T \mathbf{X}_{n,t}^R + \xi_{n,t}^R
\end{aligned}
$$

Importantly, note that the market share $S_{n,t}^R = \sum_{i \in n} s_{i,n,t}^R$ is based on the cost of sharing among all residents currently residing in neighborhood n.[48]

As the unobservable cost at the neighborhood-time level $\xi_{n,t}^R$ is allowed to be correlated with the price $p_{n,t}^A$, I instrument the short-term rental price using a measure of tourist demand

[46] For example, $\xi_{n,t}^R$ captures unobservable time trends in the cost of sharing, such as the changes in the perceived risk of hosting strangers at home as technology diffuses. It also captures unobserved costs at the neighborhood level, such as fewer house cleaners in a given neighborhood.

[47] Note that in this utility specification, I only consider random-coefficients that can be projected onto observable demographics.

[48] Note that both renters and owner-occupiers are considered as potential suppliers of home-sharing.

seasonality. Specifically, I use the number of hotel visits to the entire city of New York on the same day but lagged by seven years. Because Airbnb was essentially irrelevant before 2011,[49] it should be free of reverse causality. Unobserved structural errors in the cost of sharing should not affect the total number of New York hotel visits from seven years ago. Moreover, if one worries that the seasonality in the cost of supply remains correlated with lagged hotel demand, I also add a host of calendar-related controls, including month fixed effects, day-of-the-week fixed effects, and holiday fixed effects. What kind of variation does the price instrument leave us with? For example, it captures the impact of foreign holidays. They are persistent over time and affect hotel demand in NYC through increased tourism demand, but they are plausibly uncorrelated with the cost of home-sharing by the city residents.

The moment conditions to match include the market shares in each neighborhood n every day, together with the exogeneity of the linear shifters:

$$\forall n, t: \quad \mathbb{E}_{D_n^*}[\, s_{i,n,t}^R(\delta_{n,t}^R, \pi^R)\,] = s_{n,t}^{R,\text{o}} \tag{4.7}$$

$$\mathbb{E}[\xi^R Z] = 0 \tag{4.8}$$

where $Z = [p^{R,IV}, \mathbf{X^R}]$ includes the lagged hotel visits as the price instruments and $s_{n,t}^{R,\text{o}}$ denotes the observed market share.

To numerically estimate the supply system with over 70,000 market-share equations, I cast the problem as a minimization routine over the GMM objective. The Mathematical Programming with Equilibrium Constraints (MPEC) specification was developed in Dubé, Fox and Su (2012).

$$\begin{aligned} \min_{\delta_{n,r}^R, \alpha^R, \beta^R, \pi_\cdot^R, \eta} \quad & \eta^T W \eta \\ s.t. \quad & \forall n, t: \quad S_{n,t}^R(\delta_{n,t}^R, \pi_\cdot^R) = S_{n,t}^{R,o} \\ & \eta = Z'(\delta^R - \alpha^R p^A - \boldsymbol{\beta}^R \mathbf{X}^R) \end{aligned}$$

where W denotes the optimal weighting matrix. The primary advantage of this estimation method is the sparsity structure of the Jacobian and the Hessian. Namely, the mean utility $\delta_{n,t}^R$ only affects the equilibrium in market (n, t) and does not affect other markets.[50] It allows the optimizer to perform better numerically, especially when the number of markets is large.

[49] My data spans 2014 to 2018. Lagged by seven years, back in 2011, Airbnb had only 300 listings in NYC according to its early employees, and it has since grown over 200-fold. In 2007, Airbnb had not been founded yet. In contrast, the hotels in New York City sold over 25 million nights in 2007.
https://medium.com/@jgolden/lessons-learned-scaling-airbnb-100x-b862364fb3a7

[50] See appendix for the derivations of the relevant sparse matrices.

4.3 Parameter Estimates

Parameter Estimates from the Long-Term Rental Demand

In this section, I describe the key parameters estimated for the long-term rental demand, including both the heterogeneous preference parameters and the linear parameters. Overall, the market-share moments and the attribute covariance moments produce sensible estimates for how housing preferences vary by demographic characteristics. Then, the linear parameters are jointly estimated with the price instrument, providing measures of willingness-to-pay for each household.

Step 1 Results

By matching the covariance between housing attributes and household demographics, the estimated heterogeneous parameters reflect sensible choice patterns. Instead of showing the raw parameters, Table 4 shows the more interpretable willingness-to-pay in monthly dollars, calculated as $-\pi^L_{b,k}/\alpha^L$, where I take the price coefficient from the second step. The columns in Table 4 correspond to the vector of household demographics, whereas the rows correspond to the vector of housing and neighborhood attributes.[51]

Larger households have higher willingness-to-pay for more bedrooms. I also find strong clustering preferences for race and education. African American, Hispanic, and Asian households have significantly higher willingness-to-pay for neighborhoods with higher percentages of their own race and ethnicity. For example, an Asian household is willing to pay \$410 per month to live in a neighborhood that is one standard deviation higher in its percentage Asian.[52] In addition, educated families cluster in neighborhoods with higher proportions of educated families. The outside option is more valuable for higher income and larger families, whereas racial and ethnic minorities prefer to live in the city, all else being equal. The heterogeneity in front of the price coefficient reflects differential price sensitives, where higher-income and more educated households tend to be less price sensitive.

Step 2 Results

Since the market share moments also produce an estimate of the mean utility, Step 2 estimates the linear coefficients, including both the price coefficient and the average willingness-to-pay for each housing attribute.

[51] Neighborhood attributes are parsimoniously defined to include all attributes that are the same for all homes within that neighborhood. Specifically, it includes neighborhood demographics such as education and race. It also includes the neighborhood's location amenity as measured by average commute time. Lastly, I include an indicator for the inside option, namely, in New York City.

[52] I do not attempt to explain whether it is an Asian household's preference for its neighbor's ethnicity or for neighborhood businesses that cater to its preferences. I also do not differentiate between whether it is due to housing market preference or discrimination. These parameters simply reflect the underlying choice patterns in the data. As such, I assume whatever preference or discrimination present in the data continues to be present in the counterfactual analysis.

Table 5 column (2) shows that the price coefficient α^L is -2.04, using the instrumented estimator. The F-stat is 15.7. The use of the price instrument has a significant impact, compared to the OLS specification in column (1). To make the coefficients more interpretable, I also transform these coefficients in terms of the willingness-to-pay $(-\beta_b^L/\alpha^L)$ of the average household. Table 5 column (3) shows that the average household has a higher willingness-to-pay for more bedrooms. It also prefers pre-war structures and fewer units in the building. The average household is willing to pay \$383 for one standard deviation reduction in commuting time.

Since the entire housing stock in the market is divided into 1,050 housing types based on its neighborhood and housing attributes, the demand elasticities estimated vary by housing types. Nonetheless, the average price response to a 1% reduction in the supply of all housing types in NYC is estimated to be 1%, which implies a price elasticity of the aggregate demand of approximately 1.[53]

Parameter Estimates from the Short-Term Rental Supply

In this subsection, I describe the key parameters estimated for the supply of short-term rentals by resident hosts.

Table 6 column (4) summarizes the coefficients for the main specification. Matching the supply in each neighborhood everyday for four years results in 75,895 market-share equations to match. The linear price coefficient is 0.056, which implies an average supply elasticity of 5.96. The instrument used is the lagged hotel bookings in the city as a measure of tourist demand seasonality. The F-stat is 25.4. In comparison, in columns (1) and (2), the short-term rental prices are assumed to be exogenous, and the price coefficients obtained are biased downward significantly, suggesting the presence of unobserved costs. Besides the quadratic time trend, the main specification also includes an array of calendar-fixed effects (namely, month FE, day-of-the-week FE, and holiday FE). Column (3) shows that these fixed effects have a moderate effect on the price coefficient. The standard errors are clustered at the neighborhood level.

The average cost of home-sharing is high, estimated at \$224 per night, reflecting the fact that the overall share of hosts as a percentage of all residents is still low. In terms of cost heterogeneity, I find that lower-cost suppliers tend to be young and educated households with no children. Dividing the non-linear coefficient π_k^R by the price coefficient α^R, I find that having a college degree is associated with a reduction of \$59 per night in the cost of home-sharing. Having children in the home increases the sharing cost by \$47 per night. Being ten years younger is associated with a reduction of \$18.

[53]Recall that the outside option of the model is to reside in the bordering counties of NYC, including Hudson, Nassau, Westchester, and Bergen.

Moreover, I find a negative interaction between household income and price, which suggests that lower-income households are more price sensitive. The average supply elasticity would increase to 6.70 for those with a one-standard deviation lower in income.

5 Counterfactual Analysis

In this section, I provide estimates of the welfare and distributional impact of Airbnb in two steps. First, I discuss the welfare losses via the rent channel due to housing reallocation by absentee landlords. Second, I discuss the welfare gains via the host channel as residents act as peer suppliers. In the end, I discuss the net effects and the potential implications for the social planner.

5.1 The Distributional Impact via the Rent Channel

In this section, I conduct the counterfactual analysis to estimate the impact of Airbnb through the rent channel. Specifically, I compute the vector of counterfactual long-term rental prices p_h^L when all the housing units on Airbnb are "returned" to the long-term rental market.

Overall, I find that housing reallocation by absentee landlords aggregates to a material welfare impact for all renters in the city. Nonetheless, I find that the most significant welfare losses are suffered by renters who are higher-income, more educated, and white.

The patterns of the distributional impact are primarily driven by the geographical patterns of Airbnb penetration but compounded by the clustering preference along demographic lines. Moreover, severe housing supply restrictions result in a general equilibrium price impact of Airbnb being elevated across all NYC neighborhoods.

5.1.1 The Counterfactual Specification

Given the fully estimated long-term rental demand model, I recompute the counterfactual vector of long-term rental prices across all housing types by ensuring that housing demand equals housing supply, when the reallocated units are returned to the supply.

$$\forall h : D_h^L(p_h^{L,\text{ No Airbnb}}, p_{-h}^{L,\text{ No Airbnb}}) = S_h^F \tag{5.1}$$

Following McFadden (1978) and Small and Rosen (1981), the compensating variation can then be produced analytically when the errors are logit:

$$CV_i^L = \frac{1}{\alpha_i^L}\Big(\ln \sum_{j\in\mathcal{S}^F\setminus\mathcal{S}^A} \exp\left(V_{i,j}^L\right) - \ln \sum_{j\in\mathcal{S}^F} \exp\left(V_{i,j}^{L,\text{No Airbnb}}\right)\Big) \tag{5.2}$$

where $V_{i,j}^L = \alpha_i^L p_h^L + \beta_i^L X_h^L + \xi_h^L$ and $V_{i,j}^{L,\text{ No Airbnb}} = \alpha_i^L p_h^{L,\text{ No Airbnb}} + \beta_i^L X_h^L + \xi_h^L$ represent the non-idiosyncratic component of the utility for household i over home j of type h at the actual and the counterfactual equilibrium prices respectively.[54] $\mathcal{S}^F$ denotes the set of total physical structures available and $\mathcal{S}^A$ denotes the set of housing units observed to have been reallocated by absentee landlords.[55] Recall that a housing unit is considered to have been reallocated to the short-term rental market if it is marked "available" on Airbnb for over 180 days in 2018.[56] Across New York City, it averages to 0.68% of the rental housing stock but with remarkable heterogeneity across neighborhoods, as illustrated in Figure 4. When the counterfactual price vector is recomputed, I find an overall rent increase of 0.71%, also with substantial heterogeneity. Since the overall welfare impact depends not only on the price change in one's own housing type but also the price changes in all other housing types, I present the main results using the compensating variation measure as defined by Equation (5.2) and return to a discussion of the counterfactual equilibrium prices afterward.

In the following two sections, I first describe the distributional impact via the rent channel in terms of a variety of demographic characteristics such as household size, race and ethnicity, education, and household income. Then, I provide some characterizations to better understand the distributional patterns found. In particular, I explore (i) the role of geography, (ii) the "spillover" from one neighborhood to other neighborhoods, and (iii) the role of demographic clustering.

5.1.2 Distributional Impact by Demographic Characteristics

Overall, I find the average compensating variation via the rent channel is a loss of $138 per year, whereas the median loss is $128 per year and the standard deviation is $29. To put this in perspective, the median renter's annual household income is about $47,000. Although the magnitude translates to only 25 basis points of the annual income, the increase in equilibrium rents affects all 2.1 million households who are renters in the city. When aggregated across all renters, it amounts to a direct transfer to property owners of $200mm per year, or $2.7bn in NPV terms.

[54] Notice that the only difference between $V_{i,j}^L$ and $V_{i,j}^{L,\text{ No Airbnb}}$ is the price, while all other characteristics X_h^L and ξ_h^L are kept unchanged. Moreover, in computing the counterfactual price equilibrium, the value of the outside option is also kept unchanged.

[55] This formulation makes it explicit that the overall welfare impact of a supply squeeze on renters is comprised of two related components: the impact through an increase in the equilibrium rent and the impact through a reduction in the choice set. In particular, a consistent estimate of the compensating variation takes into account both components.

[56] This is a conservative assumption in ensuring that the unit is no longer available in the long-term rental market. There is also a concern about the selection on quality. Since the long-term rental model captures the unobserved quality for each housing type, insofar as Airbnb reallocation is more concentrated in low-quality housing types, the model fully captures this aspect. However, insofar as Airbnb reallocation might tilt towards lower-quality housing units *within* a given housing type, the model no longer captures it. Indeed, if renters in lower-quality units are more price elastic, then the counterfactual is an upper-bound of the price impact of Airbnb.

Next, to understand the differential welfare impact along demographic lines, I compute the compensating variation for each household CV_i^L and aggregate them into their respective categories. Importantly, this allows me to take into account the entire correlation matrix across demographic characteristics based on the empirical observations.

For the rest of this subsection, when I refer to "households", I restrict the analysis to all households that are renters. The welfare impact described in this subsection is also restricted to the rent channel via the long-term rental market, to be distinguished from the welfare impact via the host channel in the short-term rental market, which will be discussed in Section 5.2.

Household Size

I find that Airbnb results in larger welfare losses via the rent chaneel for smaller households. Figure 9a shows that the median welfare loss for households of size one is \$134 p.a. with a standard deviation of \$32. The median welfare loss for households of size four is \$116 p.a. Even though there remains significant variation within each household size category, the most negative welfare losses are still concentrated in small households.

In the Airbnb data, smaller housing units are disproportionately more prevalent in the short-term rental market, compared to its underlying availability. Over 80% of the entire homes listed on Airbnb (with over 180+ days availability) has fewer than two bedrooms in the NYC market. In comparison, 46% of the housing stock available for long-term rental has fewer than two bedrooms, as shown in Figure 8. This underlying pattern of Airbnb usage is one of the key drivers for why the welfare impact is more concentrated on smaller households. Nonetheless, despite the large differences in the bedroom count distribution, the differences in the welfare impact across households are much reduced. This is because a reduction in one housing type creates equilibrium price effects on all other housing types as well, which I characterize in greater detail in the next subsection.

Race and Ethnicity

I find that Airbnb results in larger welfare losses via the rent channel for white renters as compared to African American, Hispanic, or Asian renters. Figure 9b shows that the median welfare loss for white households is \$152 p.a. with a standard deviation of \$35 and a significant left tail. The median welfare loss for African American renters is \$134 p.a. The median welfare loss for Hispanic renters is \$113 p.a. The median welfare loss for Asian renters is \$127 p.a. Overall, when measured in dollar terms, the increased rent due to Airbnb reallocation hurts white renters the most, when compared to minority renters.

Figure 6 shows that there is significant clustering of housing choices along demographic lines. However, comparing it to the map of Airbnb penetration in Figure 4, one could visually discern that neighborhoods with more Airbnb rentals tend to have greater percentages of white households. Hispanic and Asian households appear to be somewhat less segregated and are generally not concentrated in regions of heightened Airbnb activities. This is also shown in Table 3, where I compute the correlation between neighborhood demographics and Airbnb reallocation, quantifying the patterns seen on the maps.[57] Although a municipal planner may choose to place different social welfare weights on different households, this analysis highlights that welfare implications depend heavily on the geographic patterns of Airbnb activity.

Education

I find that Airbnb results in larger welfare losses via the rent channel for those with more education. Figure 10a shows that the median welfare loss for renters with college degrees[58] is \$156 p.a. with a standard deviation of \$31. The median welfare loss for households without college degrees is \$120 p.a.

Similar to the previous analysis, Figure 7 shows the average education attainment across neighborhoods. A comparison with the geographic pattern of Airbnb activity in Figure 4 shows that many of the neighborhoods with higher levels of Airbnb penetration also have more educated households. Moreover, given the strong preference for those with college degrees to live in neighborhoods that have more educated households, these renters affected by Airbnb in the city center are more likely to substitute to other educated neighborhoods even if they are further away, thus "spreading" the impact to those who are more similar to them demographically. I discuss this in greater detail in Section 5.1.3.

Household Income

I find that Airbnb results in larger welfare losses via the rent channel for those with higher income. Figure 10b shows that the median welfare loss for renters in the top income quintile is \$167 p.a. with a standard deviation of \$38. The median welfare loss for renters in the bottom income quintile is \$123 p.a.

Consistent with the positive correlation between Airbnb activity and median neighborhood income shown in Table 3, the counterfactual analysis quantifies the extent to which the welfare losses are greater for higher-income renters. Besides the geography of Airbnb entry patterns, another factor is that the price coefficient α_i^L is smaller in magnitude for higher-income households, indicating that their willingness to pay for housing amenities $-\beta_i^L/\alpha_i^L$ is higher. As a

[57] It is also interesting to note that even though Airbnb activity does not appear particularly correlated with the percentage of African American households in the neighborhood, Airbnb activity tends to be high in "gentrifying neighborhoods". These neighborhoods may be identified in terms of increasing levels of education as well as declining numbers of minority households. Changes in the economic fundamentals are likely drivers for both neighborhood demographic changes and popularity among Airbnb guests. However, insofar as Airbnb reallocation further drives up rents in these areas, it can be viewed as causing the same type of changes as the ongoing gentrification process.

[58] Based on the highest educational attainment of the top earner of the household.

result, when faced with a reduction in the housing supply, the requisite compensating variation for higher-income households tends to be larger, although I show in the next subsection that this factor is secondary to geography.

5.1.3 Decomposing the Distributional Impact

In this subsection, I discuss how the estimated welfare is connected with the data and the model. In particular, I explore (a) the out-sized role that geography plays in determining the distributional impact, (b) the general equilibrium price effect because of housing supply restrictions, and (c) the role of cross-elasticity due to demographic clustering.

A. The Role of Geography

Given that I find that the higher-income and more educated renters experience larger welfare losses when housing units are reallocated to Airbnb, how much of the welfare differences across demographic groups is driven by the geographic pattern of Airbnb versus by the fact that these households tend to have a higher willingness-to-pay for housing attributes?

To disentangle these two different channels, I conduct an alternative counterfactual analysis if the housing units on Airbnb were distributed *uniformly* across geographic space and housing types. I find that less than a quarter of the distributional differences are attributable to their higher willingness-to-pay, whereas about three quarters of the differences are attributable to the geographical patterns of Airbnb activity. In the hypothetical counterfactual, if Airbnb had entered uniformly across space, Table 7 shows there would be only a $1 difference in terms of the median welfare impact between a household of size one and a household of size five. Similarly, the gap between the median impact on white renters and Hispanic renters would decrease from $39 to $7. The median welfare difference between renters with and without a college degree would decrease from $36 to $9. The gap between the top and bottom income quintiles would decrease from $43 to $9. In other words, since higher-income, and educated renters desire housing and location amenities that are particularly valuable to short-term renters, their housing units are more likely to be reallocated. As a result, they become the demographic group that is affected the most. The effect is further exacerbated as their willingness-to-pay also tends to be higher.

B. The Equilibrium Effects of Supply Restrictions

Given that the housing supply in New York City is difficult to expand, the supply restrictions exacerbate the equilibrium price response when the city experiences a supply squeeze in the long-term rental market. I first provide a theoretical derivation to decompose the overall equilibrium effect into (i) the direct effect of a supply reduction on a given housing type, (ii) the spillover effect of a supply reduction on other housing types, and (iii) the additional indirect price impact when the supply of substitute housing types is also restricted. Then, I provide some graphical intuition on the equilibrating process for illustrative purposes. Lastly, I provide numerical estimates of the breakdown between welfare transfer from the existing renters and welfare loss from the displaced renters.

First, consider the market clearing conditions for each type of housing in the long-term rental market:

$$\forall h: \quad D_h^L(p_h^L, p_{-h}^L, s^L) - s_h^L = 0 \tag{5.3}$$

where s^L denotes the entire vector of supply of each housing type.[59] Taking the total derivative of the market-clearing condition with respect to s_h and s_h' yields the following relationships:

$$\frac{dp_h^L}{ds_h^L} = \left(\frac{\partial D_h^L}{\partial p_h^L}\right)^{-1} \left(\underbrace{1}_{\substack{\text{(i). direct impact of} \\ \text{supply reduction}}} - \underbrace{\sum_{k \neq h} \frac{\partial D_h^L}{\partial p_k^L} \frac{dp_k^L}{ds_h^L}}_{\substack{\text{(iii). indirect impact from} \\ \text{price increases of other home types}}} - \underbrace{\frac{\partial D_h^L}{\partial s_h^L}}_{\substack{\text{reduction in} \\ \text{the choice set}}} \right) < 0 \tag{5.4}$$

$$\frac{dp_h^L}{ds_{h'}^L} = \left(\frac{\partial D_h^L}{\partial p_h^L}\right)^{-1} \left(0 - \underbrace{\sum_{k \neq h} \frac{\partial D_h^L}{\partial p_k^L} \frac{dp_k^L}{ds_{h'}^L}}_{\substack{\text{(ii). spillover from} \\ \text{price increases of other home types}}} - \underbrace{\frac{\partial D_h^L}{\partial s_{h'}^L}}_{\substack{\text{reduction in} \\ \text{the choice set}}} \right) < 0 \tag{5.5}$$

Equation (5.4) shows that the overall price impact from a supply reduction is a combination of the direct impact from the supply change s_h^L and the indirect impact from price changes of other housing types dp_k^L/ds_h^L. In particular, since the housing supply is *constrained* across all housing types except for the outside option, a supply reduction in a given housing type h will lead to a price increase in other housing types k because of substitution. However, this price increase in p_k^L, in turn, creates additional upward pressure on the original housing type p_h^L.

Conceptually, the housing market with severe supply restrictions is akin to a setting of imperfect competition with quantity fixing: When a competitor's price increases, the optimal response is to *increase* one's own price when the total quantity is fixed. In general equilibrium, this could lead to a more exacerbated price response than the partial equilibrium effect alone. In the housing context, even though each absentee landlord does not have market power and participates in the market competitively, the overall difficulty in expanding the housing supply acts as the quantity fixing mechanism.

[59] Given that the relevant comparative static is with respect to supply, the demand for each housing type is a function of both the vector of home prices and the entire vector of supply because of its effects on the size of the choice set.

Next, as an illustrative device, Figure 11 breaks down the equilibrating process in successive steps of partial equilibria. It shows that the general equilibrium price impact could be greater than the partial equilibrium effects, especially for neighborhoods with little direct Airbnb penetration. To illustration this, I start with the original equilibrium price vector p_h^0 that clears the original long-term rental market with supply s^F. Now, the price vector p_h^1 is allowed to respond to the supply squeeze due to Airbnb s_h^A of its own type in a partial equilibrium manner, namely by assuming all other prices remained unchanged at p_h^0. It is shown in the top left panel of Figure 11. However, this partial equilibrium price response is not sufficient to clear the market because the demand substituting into other housing types will push up their prices. Hence, each successive step m generates a new partial equilibrium price vector based on the prices from the previous step $m-1$. These are shown in the top right and the bottom left panel of Figure 11.

$$\forall h: \, D_h(p_h^m, p_{-h}^{m-1}) = s_h^F - s_h^A$$

The process continues until the equilibrium is reached for all housing types, as shown in the bottom right panel.[60] Notice that the overall quantity fixing due to supply restrictions results in a general equilibrium price impact that is much greater than the partial equilibrium alone. In neighborhoods with little Airbnb activity, the general equilibrium effects dominate.

Lastly, I quantify the welfare impact because of the changes in the equilibrium prices versus the changes in the choice set. Denote $W_i(p, \mathcal{S}) = \left(\ln \sum_{j \in \mathcal{S}} \exp V_{i,j}(p_h)\right)/\alpha_i$. The compensating variation in Eq (5.1.1) can be decomposed into the part driven by the price change and the part driven by the choice set change:

$$CV_i^L = \underbrace{\Big(W_i\big(p^L, \mathcal{S}^F \setminus \mathcal{S}^A\big) - W_i\big(p^L, \mathcal{S}^F\big)\Big)}_{\substack{\text{Welfare change due to} \\ \text{changes in the choice set}}} + \underbrace{\Big(W_i\big(p^L, \mathcal{S}^F\big) - W_i\big(p^{CF}, \mathcal{S}^F\big)\Big)}_{\substack{\text{Welfare change due to} \\ \text{changes in the equilibrium prices}}} \tag{5.6}$$

I find that the choice set reduction contributed to a welfare impact of \$58mm p.a.[61] and the increase in equilibrium housing prices contributed to a welfare impact of \$201mm p.a. The welfare transfer from renters to property owners is computed by simply summing over the price changes $\sum_h \Delta p_h^L \times (s_h^F - s_h^A)$, which amounts to \$200mm p.a. and is the first-order term (i.e. the

[60]It is important to note that there is an outside option that can accommodate the displaced renters.

[61]A reduction in the choice set itself results in a welfare decline simply because there are fewer draws of the idiosyncratic logit error $\epsilon_{i,j}^L$. This makes sense in the housing context, given that the individual-home specific utility is generally important. Even though this welfare impact does not translate into a direct transfer to the property owners, the short-term visitors do benefit from an enlarged choice set of lodging options because of Airbnb. I do not explicitly model the tourists' demand, as the focus of this paper is on the city residents. Still, I provide a back-of-the-envelope analysis at the end based on estimates from Farronato and Fradkin (2018).

dark blue rectangle in Figure 1). The welfare loss from displaced renters amounts to \$0.9mm p.a., which is the second-order term (i.e. the light blue triangle in Figure 1).

Hence, the vast majority of Airbnb's welfare impact through the long-term rental market amounts to a transfer from the remaining renters to property owners, whereas a small portion is through a reduction in the choice set, and less than 1% is due to the welfare loss of renters displaced outside of the city.

C. The Role of Demographic Clustering

In this section, I discuss that one's preference to live closer to other households with similar race and education implies that the distributional impact of Airbnb becomes compounded within demographic lines. In the NYC setting, given the overall geographical patterns of actual Airbnb activity, the clustering preference of white and educated households results in more concentrated welfare losses on white and educated renters.[62] In other words, since there is more Airbnb reallocation in centrally located and educated neighborhoods, renters affected in these areas are more likely to substitute to other neighborhoods with high levels of education even if they are located further away, thus "spreading" the price impact more heavily to other educated households.

Since the dominant pattern that emerged from the data is the preference for clustering along demographic lines, including race, ethnicity, and educational attainment, the estimated long-term rental demand model captures such rich substitution patterns. Table 8 shows a subset of the cross-elasticities, where a price increase in a given neighborhood leads to higher rates of substitution to other neighborhoods that are closer in the demographic characteristics space.

For example, Forest Hill and Jackson Heights are located in close proximity to each other in Queens, both far from the city center. Neither neighborhoods experienced much Airbnb penetration, with less than 0.5% of the housing stock being affected. However, Forest Hill has a higher proportion of educated households at 63%, whereas Jackson Heights is at 28%. On the other hand, Forest Hill is over 47% white, whereas Jackson Heights is 63% Hispanic. As a result, the model predicts that a price increase in Chelsea and Midtown, which has the highest penetration of Airbnb in the data, generates a much higher rate of substitution toward Forest Hill (0.7%) than Jackson Heights (0.1%), since both Chelsea and Forest Hill have high proportions of white and educated households, as shown in Table 8. Consequently, in the counterfactual analysis, the equilibrium price impact due to Airbnb in Forest Hill is 20% higher than in Jackson Heights.

To summarize this section, the squeeze on the long-term rental market due to Airbnb

[62] On the other hand, the clustering preference of African American and Hispanic households implies that, if more housing units were reallocated to Airbnb in predominantly minority neighborhoods, it would also generate stronger spillover to other minority renters.

results in a moderate yet material welfare transfer from renters to property owners at \$200mm p.a. or \$2.7bn in NPV terms. The median renter making \$47,000 a year loses about \$128 p.a. The general equilibrium price effect is elevated across all renters because housing supply is difficult to expand, which acts as a quantity-fixing mechanism of the market. Across the renters, I find the most significant losses are suffered by higher-income, educated, and white renters, when measured in dollar terms. The distributional differences are driven primarily by the geographical patterns of Airbnb activity and exacerbated by the preference for demographic clustering in housing choices.

5.2 The Distributional Impact via the Host Channel

In this section, I estimate the welfare derived from a resident's ability to act as a host on Airbnb. The advent of the sharing economy allows any household to participate directly in the production processes and act as a peer supplier. Overall, I find that the supplier surplus is immaterial for most households. Nonetheless, it does produce a long and heavy tail on the right, which suggests a few households with particularly low costs can benefit tremendously.

The patterns of the distributional impact are driven primarily by a household's cost to share. By observable demographic characteristics, larger surpluses accrue to households that are young, educated, and without children. Moreover, lower-income households have a more elastic supply, which suggests that they benefit more at peak times and locations.

5.2.1 The Counterfactual Specification

To evaluate the welfare gains from direct home-sharing, I perform a counterfactual analysis where the option to host on Airbnb is no longer available to the residents. The compensating variation of household i residing in neighborhood n is computed as follows

$$CV_i^R = \frac{1}{\alpha_i^R} \sum_t \ln\left(1 + \exp(V_{i,t}^R)\right) \tag{5.7}$$

where $V_{i,t}^R = \alpha_i^R p_{n,t}^A + \beta_i^R X_{n,t}^R + \xi_{n,t}^R$ and is summed over the course of the year.[63] Note that this specification calculates the *ex-ante* expected welfare from sharing before the actual idiosyncratic component $\epsilon_{i,t}^R$ is realized. Consequently, the distribution of the expected surplus is different from the distribution of the realized surplus, as the latter includes the additional variance due to the variance in the error term.

Overall, the distribution of the supplier surplus is centered close to zero but with a heavy right tail, suggesting that the bulk of the benefits is accrued to a concentrated few. Figure 12

[63] I use 2018 as the base year over which the surplus is computed.

shows the surplus distribution of all renters in the city,[64] where the median supplier surplus is only \$0.4 p.a. At the 75th percentile, the supplier surplus remains immaterial at \$5.9 p.a. However, the expected surplus on the very right tail above the 99th percentile amounts to \$307 p.a. When integrated over all renters, the total surplus produced by direct home-sharing amounts to \$23mm a year, or \$300mm in NPV terms.[65]

5.2.2 Supplier Surplus by Demographic Characteristics

One of the key benefits of estimating a random-coefficient supply system is its ability to analyze the supplier surplus by observable demographic characteristics, including income, education, age, and family structure. I discuss how each affects the surplus through one's cost of sharing as well as one's price sensitivity. Since the surplus is immaterial for the majority of the renters, I discuss patterns for both the typical renter and the right tail.

Household Income

Even though the home-sharing surplus is immaterial for the majority of the households, higher-income households still expect to have a larger surplus on average. Meanwhile, the lowest-income group enjoys the largest benefits in the tail.

Figure 13 shows that the median host surplus for renters in the top income quintile is \$5.2 p.a., and the median host surplus for renters in the bottom income quintile is \$0.1 p.a. Other non-extreme percentiles (e.g. 75th or 90th) of supplier surplus are also generally higher for higher-income households. By contrast, the average surplus above the 99th percentile is significant at \$232 p.a. for the top income quintile. But it is even higher at \$454 p.a. for the bottom income quintile.

The difference between the median and the tail outcome is a result of two countervailing forces. Intuitively, the decision to share is based on a comparison of the market price $p^A_{n,t}$ and the cost to share $|\beta^R_i|/\alpha^R_i$. On the one hand, higher household income (as well as other correlates of income such as education) results in a lower cost to share with a smaller $|\beta^R_i|$ estimated from the model. On the other hand, higher household income also lowers one's price sensitivity α^R_i, suggesting that the sharing income may not be as valuable. In other words, lower-income households are more likely to find the hassle of hosting visitors at home worth the money.

[64]In this section, I focus on the supplier surplus of renters in New York City, to make it directly comparable to the previous section. Nonetheless, the supplier surpluses of all residents, including both renters and owner-occupiers, are estimated and are included in the aggregate measures when appropriate.

[65]In comparison, conditioning on having made a room available on Airbnb in 2018 (totaling 24,100 listings, or 0.8% of the occupied housing units), the median revenue obtained by such resident hosts is \$2,484 in 2018, totaling \$137mm. Note that this represents only a fraction of the total Airbnb revenue in NYC, as the majority of its revenue is earned by housing units operated by absentee landlords.

Relatedly, as there is significant demand seasonality in the short-term rental market, the sharing income during peak demand times could be particularly valuable for lower-income households

$$\frac{\partial CV_{i,t}^R}{\partial p_{n,t}^A} = \frac{\exp\left(\alpha_i^R p_{n,t}^A + \beta_i^R X_{n,t}^R + \xi_{n,t}^R\right)}{1 + \exp\left(\alpha_i^R p_{n,t}^A + \beta_i^R X_{n,t}^R + \xi_{n,t}^R\right)} \quad \Rightarrow \quad \frac{\partial^2 CV_{i,t}^R}{\partial p_{n,t}^A \partial \alpha_i^R} > 0$$

The cross-partial derivative of the compensating variation with respect to the market price $p_{n,t}^A$ and the price coefficient α_i^R is positive, where α_i^R is decreasing in household income. Given the model parameters, the supply elasticity for the top income quintile is 5.0, whereas the supply elasticity for the bottom income quintile is 7.4.

Overall, the anecdotal narrative that home-sharing could be beneficial for low-income families is borne out only in a limited sense: Conditional on being in the right tail of the surplus distribution, there are more low-income families, and they are more willing to take advantage of the peak demand times. However, for the typical family, regardless of income level, their expected benefit from home-sharing is immaterial.

Other Demographic Characteristics

I find that households that are younger and have no children obtain greater supplier surpluses as resident hosts. For the youngest households with no children, the average surplus is expected to be $80 p.a. For households with children, the average surplus is immaterial across all age groups.[66] Since age and family structure only enter the cost term β_i^R in the model, they are not in front of the price coefficient α_i^R. As a result, both the average and the tail of the surplus distribution load heavily on young households with no children.

In terms of education, I find that educated households accrue greater supplier surplus as resident hosts. On average, the expected surplus is $36 p.a. for educated households and immaterial for households without a college degree.[67] Again, both the average and the tail of the surplus distribution load heavily on educated households.

Hence, even a reasonably simple supply system captures important cost heterogeneity of peer suppliers, where larger benefits accrue to those who are young, educated, and without children. In addition, it captures particularly rich dynamics with respect to household income, where a few low-income households could benefit a lot from the home-sharing platform. Nonetheless, for the typical household, the cost to share one's home with visitors is high; thus,

[66] There are likely richer dynamics with respect to age-related events. However, given that the current model is estimated on aggregate data, it only captures a number of key demographic characteristics.

[67] The model does not necessarily explain why more education is likely associated with a lower cost of sharing, although one might speculate on reasons related to financial sophistication, willingness to adopt new technology, and one's ability to manage online businesses.

the surplus remains immaterial.

5.3 The Net Impact on Renters

In this section, I estimate the net welfare impact of Airbnb on renters by combining the welfare losses via the rent channel and the supplier surplus via the host channel described in the two previous sections. The net effect for the median renter is negative as the rent channel dominates, but the right tail is long as a few households benefit heavily from hosting. Nonetheless, on average, the losses remain more significant for educated, higher-income, and white renters.

In addition, the net welfare is divergent between the median and the tail in terms of its spatial distribution, where neighborhoods with the largest median losses also tend to be areas with the largest gains in the tail. I conclude with a discussion of how the social planner's objective may differ from that of the residents.

5.3.1 The Distribution of the Net Welfare Impact

Since the losses from the rent channel are diffused and the gains from the host channel are concentrated, the net welfare impact for the median renter is a loss of \$125 p.a, as shown in Figure 14. At the 75th percentile, the net welfare impact is -\$113 p.a. In fact, the net welfare for over 97% of the renters is negative. Nonetheless, the long right tail in the distribution of host gains results in a similarly long tail in the net welfare impact, averaging to \$164 p.a. above the 99th percentile.[68] In the remainder of the section, I decompose the net welfare impact by different demographic characteristics (income, race, education) and by neighborhood geographic location.

Net Welfare by Household Income

Overall, higher-income renters experience larger net losses on average and experience smaller net gains in the tail. As such, when restricted to just renters, it does not seem that Airbnb exacerbates income inequality. However, one needs to be more cautious of the overall impact beyond just renters, because unregulated Airbnb will likely worsen the wealth inequality due to the large welfare transfer from renters to property owners.

Table 9 shows that renters in the top income quintile lose \$146 p.a. on average because of Airbnb, driven primarily by the welfare loss through the rent channel at \$169, as the gain from

[68] As discussed before, the ex-post realized gain for the top percentile outcome is likely more extreme than the ex-ante expected surplus. Hence, the tail statistics should be interpreted accordingly as the ex-ante expectations before the idiosyncratic error terms are realized. I also conduct a robustness analysis with alternative assumptions on the auto-correlation of the error term, where the results are also qualitatively similar.

the host channel is only $24. For renters in the bottom income quintile, the average net welfare is -$114, where both the losses from the rent channel and the gains from the host channel are less pronounced when compared with higher-income renters.

In the right tail, the increased price sensitivity of lower-income households results in greater host surpluses, especially if the household lives in areas with heavy short-term rental demand, as discussed in the previous section. In other words, for a small proportion of households, especially low-income ones, when their cost of sharing is low, they can not only make up for the higher rents but also obtain significant surpluses from becoming a host. As higher-income households are less likely to find the hassle of home-sharing worthwhile, their gains in the tail are smaller. Therefore, higher-income renters fare worse both on average and in the tail.

Net Welfare by Education

In terms of education, the median net welfare impact is worse for more educated renters. Table 10 shows that the median is -$136 p.a. for renters with a college degree and -$120 for those without a college degree. As education is associated with a lower cost of home-sharing, I find that educated renters are more likely to be in the right tail of the host surplus distribution than those without a college degree. The divergence of the median and the tail by education is a result of the interaction between the demand and cost parameters in the short-term rental market.

Net Welfare by Race and Ethnicity

When evaluating the net welfare impact along race and ethnicity lines, I find that the median white renter loses more than the median African American or Hispanic renter. This difference between the group medians is primarily driven by the differential welfare losses via the rent channel, where the proportion of housing units reallocated to Airbnb is higher in neighborhoods with more white and educated renters. The net welfare impact on the median white renter is -$152 p.a. In comparison, the net impact on the median African American renter is -$134 p.a., and for the median Hispanic renter, it is -$113 p.a.

Even though race and ethnicity itself do not enter into the cost of home-sharing directly, there remains an induced distribution in terms of the host surplus because of the correlation among race, education, and income. As a result, I find more white households in the right tail of the welfare distribution than African American or Hispanic households.[69] White households

[69] Although it is not unreasonable to start with a model where race or ethnicity per se do not affect the cost of home-sharing, once controlling for other demographic characteristics, it is possible that African American and other minority hosts might face discrimination from the demand side and thus not be able to obtain as much surplus as peer suppliers. In this case, it would further widen the racial gap found here.

in the tail make \$326 p.a. from hosting, whereas Black and Hispanic households make \$227 p.a. and \$232 p.a., respectively.

Net Welfare by Geography

In addition to decomposing the net welfare impact by demographic characteristics, it is also instructive to decompose the results by neighborhood location, especially since the short-term rental demand varies substantially across the geographic space. Two trends emerge in the spatial distribution of the net welfare: (i) There is a divergence between the experience of the median and the tail renter. (ii) The variation in terms of how much host gains could offset rent increases is driven by the neighborhood's cost of home-sharing, which favors centrally located low-income neighborhoods.

The rightmost panels in Figure 15 and Figure 16 shows the net welfare impact on the median renter and the right tail respectively. Interestingly, neighborhoods that experience heavier losses for the median renter also tend to experience large net gains in the right tail. For example, the net welfare impact for the median renter in Chelsea, which has high levels of Airbnb penetration, is -\$146 p.a., while the tail of its host surplus is above \$600 p.a. The irony of the divergence is explained by the overall spatial patterns of short-term rental demand: High short-term demand for the neighborhood drives more reallocation of the housing units away from the long-term rental market, thereby raising rents for everyone in the neighborhood. At the same time, high short-term rental demand also increases the prices that resident hosts can benefit from when they share their homes. However, as low-cost resident hosts are relatively few in number, their large gains affect only the tail of the distribution, not the median.

Although the short-term rental demand drives both the rent increases and the host surpluses, the variation in their difference is affected by the cost of home-sharing, where there are more low-cost hosts in low-income neighborhoods. For instance, the median household income among renters in Bushwick is only 60% of the median household income of the neighboring Park Slope and Williamsburg. As such, residents there are more likely to find home-sharing worthwhile. Despite being a relatively lower-income area, Bushwick is still conveniently located in Brooklyn and attracts short-term visitors. As a result, the surplus from home-sharing is relatively more valuable to its residents.[70]

5.4 Implications for the Planner

In this section, I discuss the aggregate welfare impact on all participants of the long-term and short-term rental market by combining model estimates from the previous sections. Despite an increase in the overall efficiency, the crucial trade-off is that the advent of home-sharing tends

[70] For example, as measured in terms of a larger residual when regressing the host gains on rent increases at the neighborhood level. Also, it can be visually inspected by comparing the two top panels of Figure 15.

to benefit those who already own property. In the presence of severe housing market supply restrictions, the welfare transfer from renters to owners becomes substantial. In a city where the majority of the residents are renters, unregulated Airbnb will likely hurt the median person.

The top panel of Table 11 summarizes the aggregate impact on renters. On the one hand, the supply model estimates a surplus of \$23mm p.a. from home-sharing. On the other hand, the transfer from renters to absentee landlords amounts to \$200mm p.a., and the welfare loss from displaced renters is about \$1mm p.a. Therefore, the net impact on renters is a loss of \$178mm per year, or \$2.4bn in NPV terms. Although the total surplus from the host channel appears much smaller in magnitude compared to the loss from the rent channel, the relevant comparison for the social planner is different from that of the renter: The welfare gains obtained through the host channel are *true* economic benefits due to the innovation, whereas the welfare loss via the rent channel is not actually lost, but primarily a *transfer* to existing property owners.

If I include all participants of the housing market, namely renters, owner-occupiers, and absentee landlords, the aggregate welfare impact then becomes positive, as illustrated in the second panel of Table 11. Assuming absentee landlords choose to operate between the long-term and the short-term market rationally, they also accrue surplus from hosting when they operate on Airbnb. A back-of-the-envelope estimation based on their actual Airbnb revenue less the forgone rents results in a surplus of \$46mm per year.[71] As owner-occupiers can be viewed as paying rent to themselves, the increase in property value is then mostly offset by the increase in the equivalent user-cost of their home.[72] Even though the distributional analysis in the previous section does not suggest that Airbnb exacerbates income inequality *among* renters, the presence of severe housing supply restrictions implies that the reallocation channel opened up by Airbnb will likely increase the wealth inequality between property owners and renters.[73]

Moreover, there is an additional net surplus for tourists. In the third panel of Tabel 11, using estimates from Farronato and Fradkin (2018), a back-of-the-envelope approximation produces a substantial gain from the consumer surplus of short-term visitors, net of hotel losses. As such, the total efficiency gain from the social planner's perspective is strictly positive and substantial.

Despite the overall increase in efficiency, the presence of housing supply restrictions results in a net welfare loss for over 97% of the renters. As 67% of the housing units are renter-occupied, the median person in the welfare distribution is a renter. The estimated net welfare on the

[71] Even though this is not precisely estimated based on an actual cost model of absentee landlords operating on Airbnb, the fact that it is much greater than the welfare loss of the displaced renters (\$1mm p.a.) is what makes the net impact on all housing market participants positive.

[72] For owner-occupiers, since they also have the option to leave the city, this suggests they are better off.

[73] In other words, had housing supply been elastic, one would expect a muted price response in the long-term rental market, a bigger reduction in price in the short-term rental market, and the supplier surplus to accrue to those hosts who have low costs of production.

median resident is a loss of \$114 p.a.[74] Consequently, a simple voting model favors regulations to restrict Airbnb, especially restricting the reallocation of housing units by absentee landlords. In practice, it is consistent with the regulatory actions taken in New York. If such restrictions were fully enforced, a large portion of the tourist gains would be erased, whereas the rent transfer would be reversed. However, the analysis from the previous section suggests that enforcing Airbnb restrictions will also likely lead to a bigger reduction of housing costs for higher-income, educated, and white renters.

In other words, even though restricting Airbnb seems to be the most immediate regulatory action that can improve the median voter's outcome, the analysis here suggests that the true underlying challenge remains the inability for the city to expand housing supply. The entry of Airbnb provides a channel for the existing space to be used by the highest value bidder, but the total quantity restriction raises rents for everyone. Therefore, a reasonable alternative regulatory approach is to allow housing supply to expand more easily.

5.4.1 Model Limitation and Extensions

The structural model in this paper delivers reasonable parameter estimates as well as rich counterfactuals. However, I would like to point out its inherent limitations and the extensions that enrich the existing framework.

Both the long-term rental demand and the short-term rental supply model are static, which ignores transition dynamics. In the long-term rental model, I have assumed away moving costs. As such, the welfare impact found in a frictionless market can only be viewed as an approximation to the actual effects at best. Nonetheless, if moving costs are incorporated, the extent of welfare impact faced by renters in the most affected neighborhood is likely larger, which further exacerbates the losses suffered by higher-income, educated, and white renters. In the short-term rental market, there is an adoption process of the technology, where resident hosts exert a one-time effort to list on Airbnb and incur ongoing expenses to host guests. The static model abstracts away such adoption costs, assuming that the ongoing expenses are dominant.

There are a number of potentially useful extensions. First, the current model focuses on the increased housing costs faced by renters, since the owner-occupiers are always weakly better off. However, it is useful to characterize the gains accrued to the "displaced" owner-occupiers who sell their homes at an increased price and leave the city, benefiting even more than the owner-occupiers who remain in the city.

Second, I have yet to model the joint housing choice problem when households take into account their expected short-term rental surpluses as they make long-term rental choices. Such

[74] This assumes renters in the city do not own residential properties in the city, which is likely true for the vast majority of the renters.

a joint decision model predicts that households with low costs of sharing (i.e. young, educated families with no children) will move further toward neighborhoods that are popular among tourists. Moreover, the expected short-term rental value will shift out the long-term demand curve, which results in even higher housing prices in equilibrium.

Last, the existing rent control and rent stabilization regulations could be modeled explicitly. Insofar as the current rental control law creates a mismatch in the housing market, the ability to rent part of a home out on Airbnb reverses its distortive effect.[75] In the current model, it is estimated as part of the host gains and captured only in the unobserved cost component at the neighborhood level. Meanwhile, as rent-stabilized units have limited ability to adjust rent, the impact of Airbnb on prices becomes further concentrated on the remaining market-price units, which tend to have higher-income tenants.

6 Conclusion

In this paper, I estimate the impact of the sharing economy on the highly contentious New York City housing market.

Such sharing technology operates on two fronts: the reallocation of resources and the increased utilization of resources. In a supply-constrained market, the reallocation of housing to Airbnb leads to an increase in equilibrium rents across all housing units in the city, not just for the specific units removed. It results in a widespread loss for all renters, aggregating to a transfer of \$200mm p.a. to property owners. Moreover, the heterogeneity allowed in my structural model shows that more significant losses are shouldered by renters who are higher-income, more educated, and white as they tend to desire housing and neighborhood amenities that are highly desirable to short-term visitors as well. The utilization channel allows residents to provide short-term rental services in their existing homes. The estimated supply model suggests the cost of home-sharing remains high for most people, as evidenced by the fact that less than 1% of the residents become hosts. As a result, the median host gain is immaterial, aggregating to \$23mm p.a. across the city. Nonetheless, a small fraction of the households with particularly low-cost of sharing obtains substantial host surpluses, including a few enterprising low-income families taking advantage of peak short-term rental demand.

As Mayor de Blasio pushes for stricter enforcement of short-term rental regulations in New York City,[76] it is partially consistent with the goal of alleviating the housing affordability crisis, but it does not address the more fundamental problems created by the housing supply

[75] In fact, the presence of existing housing-market regulations tends to generate even more regulations, as Airbnb would otherwise provide a channel for regulatory arbitrage. For example, the short-term rental regulation in Los Angeles specifically does not allow units covered by the Rent Stabilization Ordinance to be on Airbnb.

[76] Under the current Multiple Dwelling Law in New York, short-term rentals of Class A properties are not allowed unless its permanent resident is present. If fully enforced, it effectively rules out housing reallocation by absentee landlords, but still helps to protect some of the host gains.

constraints. Banning the reallocation of housing to Airbnb likely reduces the housing costs for all renters, and especially help higher-income and educated renters. However, such ban will be at the expense of significant losses to both existing property owners and potential tourists who never arrive. It also creates additional incentives to evade regulations (Jia and Wagman, 2018). Such near-term regulation against Airbnb masks the underlying challenge, namely the detrimental effects of housing supply restrictions, which afflict not only New York City,[77] but also many other productive cities across the United States.[78] Even without Airbnb, these housing supply restrictions continue to produce significant economic distortions because of the inefficient location choices made by workers and firms (Hsieh and Moretti, 2019).

More broadly, the welfare impact of the sharing economy, as well as many other financial market innovations, can arise from a purely technological aspect as well as a regulatory aspect. The reallocation of housing units due to Airbnb serves as a form of regulatory arbitrage which reduces the price wedge created by the pre-existing allocation of space between the long-term rental market and the short-term rental market. It also responds to regulations that increase the cost of operating in the long-term rental market (e.g., rent control, tenant protection laws) by allowing property owners to generate cash flow in an alternative market. In other contexts, the development of novel financial products often serves as a form of regulatory arbitrage, fueling the growth of shadow banking or shadow insurance (Buchak, Matvos, Piskorski and Seru, 2018; Koijen and Yogo, 2016).

The utilization aspect of the sharing economy is a more welcoming technological feature, as it allows enterprising residents, including low-income ones, to engage in business activities that would otherwise be costly to start. By aggregating demand and verifying payments, as well as having substantial network effects, these peer-to-peer platforms reduce the barriers to entry for many enterprising individuals. The growth and behavior of a whole class of platform entrepreneurs is an exciting avenue for future research.

[77]Although physical geography plays a role as Manhattan is a peninsula, the primary source of housing supply restrictions remains regulatory, ranging from restrictive zoning to onerous floor area ratios.

[78]In fact, it is also consistent with the observation that Airbnb tends to face more legal troubles in cities with more housing constraints.

References

Alonso, William. 1964. "Location and Land Use." *Harvard University Press.*

Autor, David, Christopher Palmer, and Parag Pathak. 2017. "Gentrification and the Amenity Value of Crime Reductions: Evidence from Rent Deregulation." *NBER Working Paper.*

Autor, David H., Christopher Palmer, and Parag A. Pathak. 2014. "Housing Market Spillovers: Evidence from the End of Rent Control in Cambridge Massachusetts." *Journal of Political Economy*, *122* (3): 661–717.

Banzhaf, H. Spencer and Randall P. Walsh. 2008. "Do People Vote with Their Feet? An Empirical Test of Tiebout's Mechanism." *American Economic Review*, *98* (3): 843–863.

Barron, Kyle, Edward Kung, and Davide Proserpio. 2018. "The Sharing Economy and Housing Affordability : Evidence from Airbnb."

Bayer, Patrick and Christopher Timmins. 2005. "On the Equilibrium Properties of Locational Sorting Models." *Journal of Urban Economics*, *57*, 462–477.

Bayer, Patrick and Christopher Timmins. 2007. "Estimating Equilibrium Models for Sorting Across Destinations." *The Economic Journal*, *117*, 353–374.

Bayer, Patrick, Fernando Ferreira, and Robert Mcmillan. 2007. "A Unified Framework for Measuring Preferences for Schools and Neighborhoods." *Journal of Political Economy*, *115* (4).

Bayer, Patrick, Nathaniel Keohane, and Christopher Timmins. 2009. "Migration and Hedonic Valuation: The Case of Air Quality." *Journal of Environmental Economics and Management*, *58* (1): 1–14.

Benetton, Matteo. 2018. "Leverage Regulation and Market Structure: An Empirical Model of the UK Mortgage Market." *SSRN Working Paper Series.*

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica*, *63* (4): 841–890.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1999. "Voluntary Export Restraints on Automobiles: Evaluating a Strategic Trade Policy." *American Economic Review*, *89* (3): 400–430.

Berry, Steven, James Levinsohn, and Ariel Pakes. 2004. "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market." *Journal of Political Economy*, *112* (1): 68–105.

Black, Sandra E. 1999. "Do Better Schools Matter? Parental Valuation of Elementary Education." *Quarterly Journal of Economics*, *114* (2): 577–599.

Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru. 2018. "Fintech, Regulatory Arbitrage, and the Rise of Shadow Banks." *Journal of Financial Economics*, *130* (3): 453–483.

Calder-Wang, Sophie. 2019. "No Land Like Disneyland: The Impact of Airbnb Restriction on Home Prices." *Working Paper.*

Campbell, John, Stefano Giglio, and Parag Pathak. 2011. "Forced Sales and House Prices." *American Economic Review*, *53* (9): 1689–1699.

Chamberlain, Gary. 1987. "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions." *Journal of Econometrics*, *34* (3): 305–334.

Cuesta, José Ignacio and Alberto Sepulveda. 2019. "Price Regulation in Credit Markets: A Trade-Off between Consumer Protection and Credit Access." *Working Paper.*

Davis, Lucas W. 2011. "The Effect of Power Plants on Local Housing Values and Rents." *Review of Economics and Statistics*, *93* (4): 1391–1402.

Dubé, Jean-Pierre, Jeremy T. Fox, and Che-Lin Su. 2012. "Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand Estimation." *Econometrica*, *80* (5): 2231–2267.

Edelman, Benjamin, Michael Luca, and Dan Svirsky. 2017. "Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment." *American Economic Journal: Applied Economics*, pp. 1–34.

Epple, Dennis. 1987. "Hedonic Prices and Implicit Markets : Estimating Demand and Supply Functions for Differentiated Products." *Journal of Political Economy*, *95* (1): 59–80.

Farronato, Chiara and Andrey Fradkin. 2018. "The Welfare Effects of Peer Entry in the Accommodation Market: The Case of Airbnb." *NBER Working Paper.*

Ferreyra, Maria Marta. 2007. "Estimating the Effects of Private School Vouchers in Multidistrict Economies." *American Economic Review*, *97* (3): 789–817.

Filippas, Apostolos, John J. Horton, and Richard J. Zeckhauser. 2019. "Owning, Using and Renting: Some Simple Economics of the 'Sharing Economy'." *Working Paper.*

Galiani, Sebastian, Alvin Murphy, and Juan Pantano. 2015. "Estimating Neighborhood Choice Models: Lessons from a Housing Assistance Experiment." *American Economic Review*, *105* (11): 3385–3415.

Ganong, Peter and Daniel Shoag. 2017. "Why Has Regional Income Convergence in the U. S. Declined?" *Journal of Urban Economics*, *102*, 76–90.

Glaeser, Edward and Joseph Gyourko. 2018. "The Economic Implications of Housing Supply." *Journal of Economic Perspectives*, *32* (1): 3–30.

Greenstone, Michael and Justin Gallagher. 2008. "Does Hazardous Waste Matter? Evidence from the Housing Market and the Superfund Program." *Quarterly Journal of Economics*, (August): 951–1003.

Gyourko, Joseph. 2009. "Housing Supply." *Annual Review of Economics*, *1*, 295–318.

Gyourko, Joseph and Raven Molloy. 2015. "Regulation and Housing Supply." *Handbook of Regional and Urban Economics*, *5*, 1289–1337.

Higgins, Sean. 2019. "Financial Technology Adoption." *Working Paper.*

Horn, Keren and Mark Merante. 2017. "Is Home Sharing Driving up Rents? Evidence from Airbnb in Boston." *Journal of Housing Economics*, *38*, 14–24.

Hsieh, Chang-Tai and Enrico Moretti. 2019. "Housing Constraints and Spatial Misallocation." *American Economic Journal: Macroeconomics.*

Jaffe, Sonia, Peter Coles, Steven Levitt, and Igor Popov. 2019. "Quality Externalities on Platforms: The Case of Airbnb." *Working Paper.*

Jia, Jian and Liad Wagman. 2018. "Platform, Anonymity, and Illegal Actors: Evidence of Whac-a-Mole Enforcement From Airbnb." *SSRN Electronic Journal.*

Klaiber, H. Allen and Daniel J. Phaneuf. 2010. "Valuing Open Space in a Residential Sorting Model of the Twin Cities." *Journal of Environmental Economics and Management*, *60* (2): 57–77.

Koijen, Ralph S. J. and Motohiro Yogo. 2016. "Shadow Insurance." *Econometrica*, *84* (3): 1265–1287.

McFadden, Daniel. 1978. "Modeling the Choice of Residential Location." *Spatial Interaction Theory and Planning Models*, pp. 72–77.

Mills, Edwin S. 1967. "An Aggregative Model of Resource Allocation in a Metropolitan Area." *American Economic Review: Papers & Proceedings*, *57* (2): 197–210.

Muth, Richard F. 1969. "Cities and Housing: the Spatial Pattern of Urban Residential Land Use." *University of Chicago Press*, (18): 200.

Nelson, Scott T. 2019. "Private Information and Price Regulation in the US Credit Card Market." *Working Paper.*

Nevo, Aviv. 2000. "Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry." *The RAND Journal of Economics*, *31* (3): 395.

Nevo, Aviv. 2001. "Measuring Market Power in the Ready-to-Eat Cereal Industry." *Econometrica*, *69* (2): 307–342.

Reynaert, Mathias and Frank Verboven. 2014. "Improving the Performance of Random Coefficients Demand Models: The Role of Optimal Instruments." *Journal of Econometrics*, *179* (1): 83–98.

Robles-Garcia, Claudia. 2018. "Competition and Incentives in Mortgage Markets: The Role of Brokers." *Working Paper.*

Rosen, Sherwin. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy*, *82* (1): 34–55.

Saiz, Albert. 2010. "The Geographic Determinants of Housing Supply." *Quarterly Journal of Economics*, *125* (3): 1253–1296.

Small, Kenneth A and Harvey S Rosen. 1981. "Applied Welfare Economics with Discrete Choice Models." *Econometrica*, *49* (1): 105–130.

Timmins, Christopher. 2007. "If You Cannot Take the Heat, Get out of the Cerrado... Recovering the Equilibrium Amenity Cost of Nonmarginal Climate Change in Brazil." *Journal of Regional Science*, *47* (1): 1–25.

Tra, Constant I. 2010. "A Discrete Choice Equilibrium Approach to Valuing Large Environmental Changes." *Journal of Public Economics*, *94* (1-2): 183–196.

Valentin, Maxence. 2019. "The Effects of Regulating the Housing Short-Term Rental Market: Evidence From New Orleans." *Working Paper*, *16802*, 1–38.

Wachsmuth, David and Alexander Weisler. 2018. "Airbnb and the Rent Gap: Gentrification through the Sharing Economy." *Environment and Planning A: Economy and Space*, *50* (6): 1147–1170.

Wong, Maisy. 2019. "A Tractable Framework To Relate Marginal Willingness-To-Pay in Hedonic and Discrete Choice Models." *Journal of Housing Economics.*

7 Figures

Figure 1: An Illustrative Model of an Integrated Housing Market

Panels A and B illustrate the market equilibrium when Airbnb allows for a more efficient allocation of housing units *between* the long-term and the short-term rental market. Panels C and D illustrate the market equilibrium when Airbnb also allows for increased utilization of existing homes already occupied by long-term tenants.

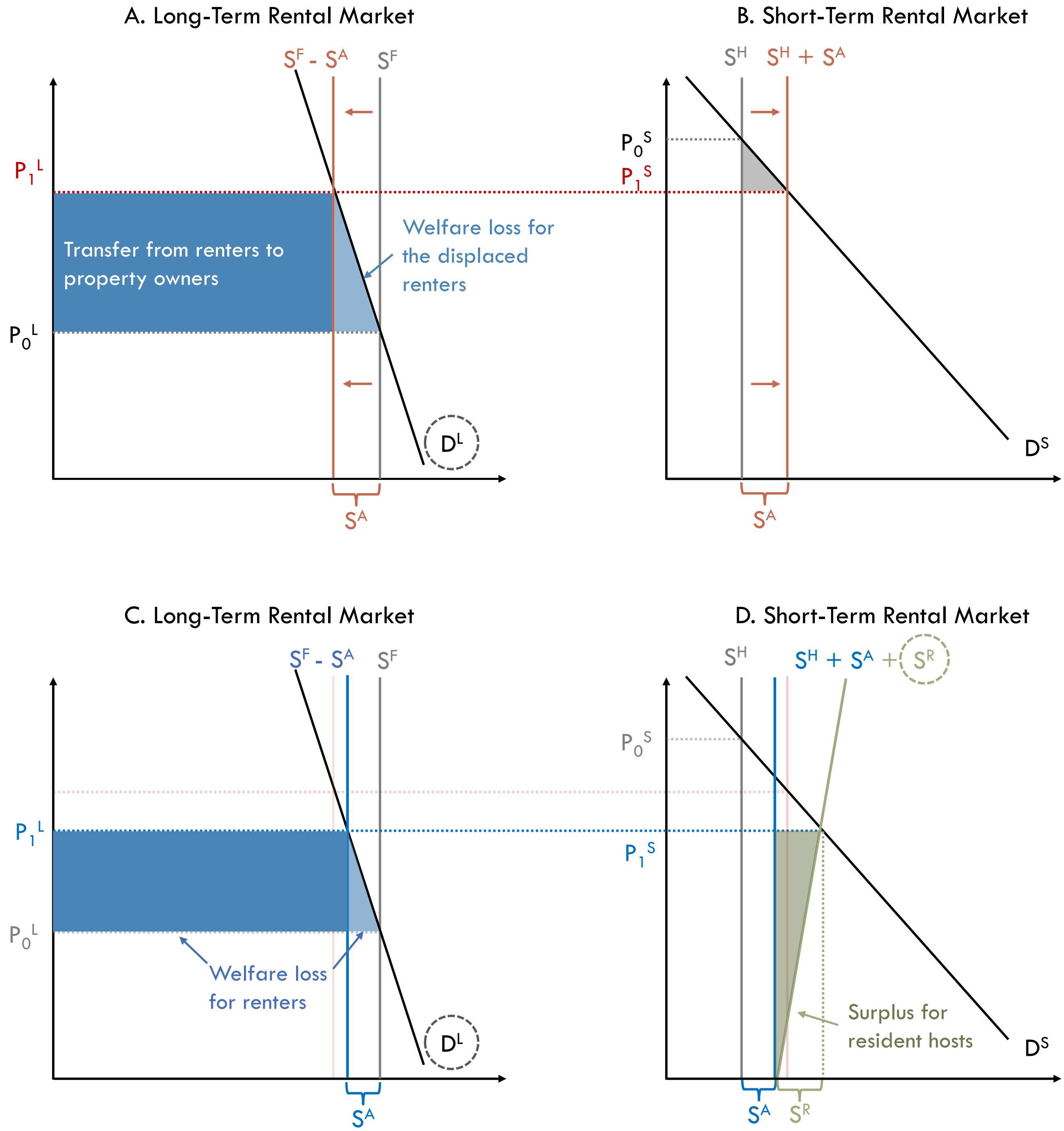
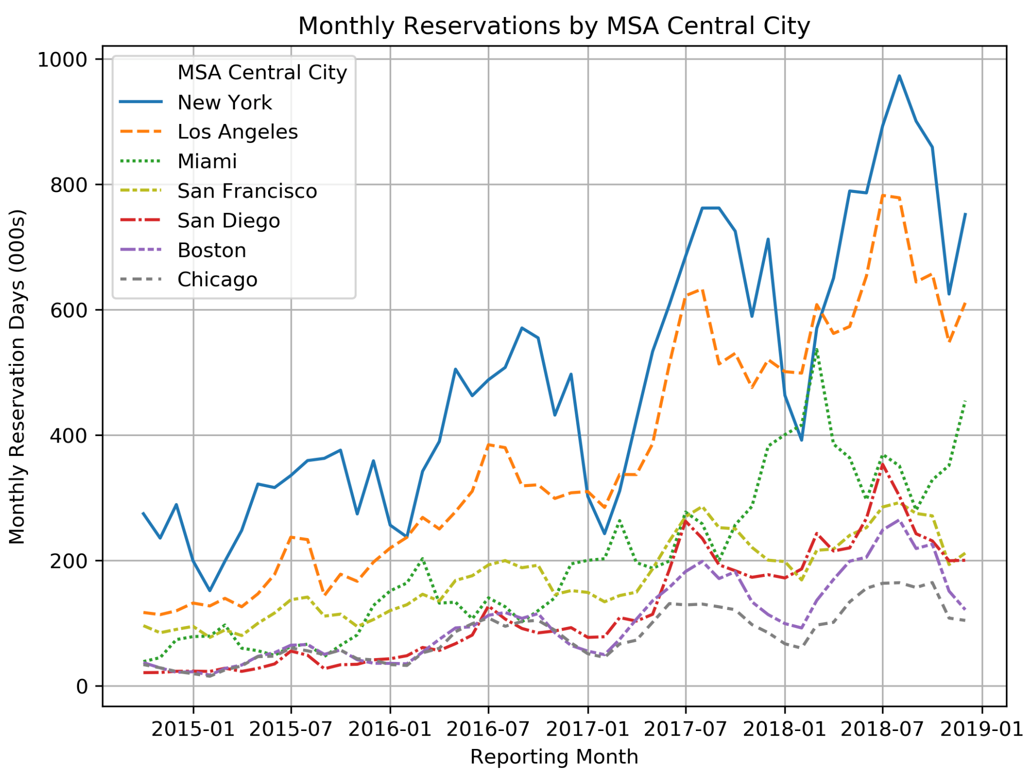


Figure 2: Growth of Airbnb across the U.S.

The time series plot shows the rapid growth in the number of monthly reservations across select MSAs in the United States. New York is the largest metropolitan market for Airbnb in the United States. Between 2015 and 2018, the number of reservations quadrupled.



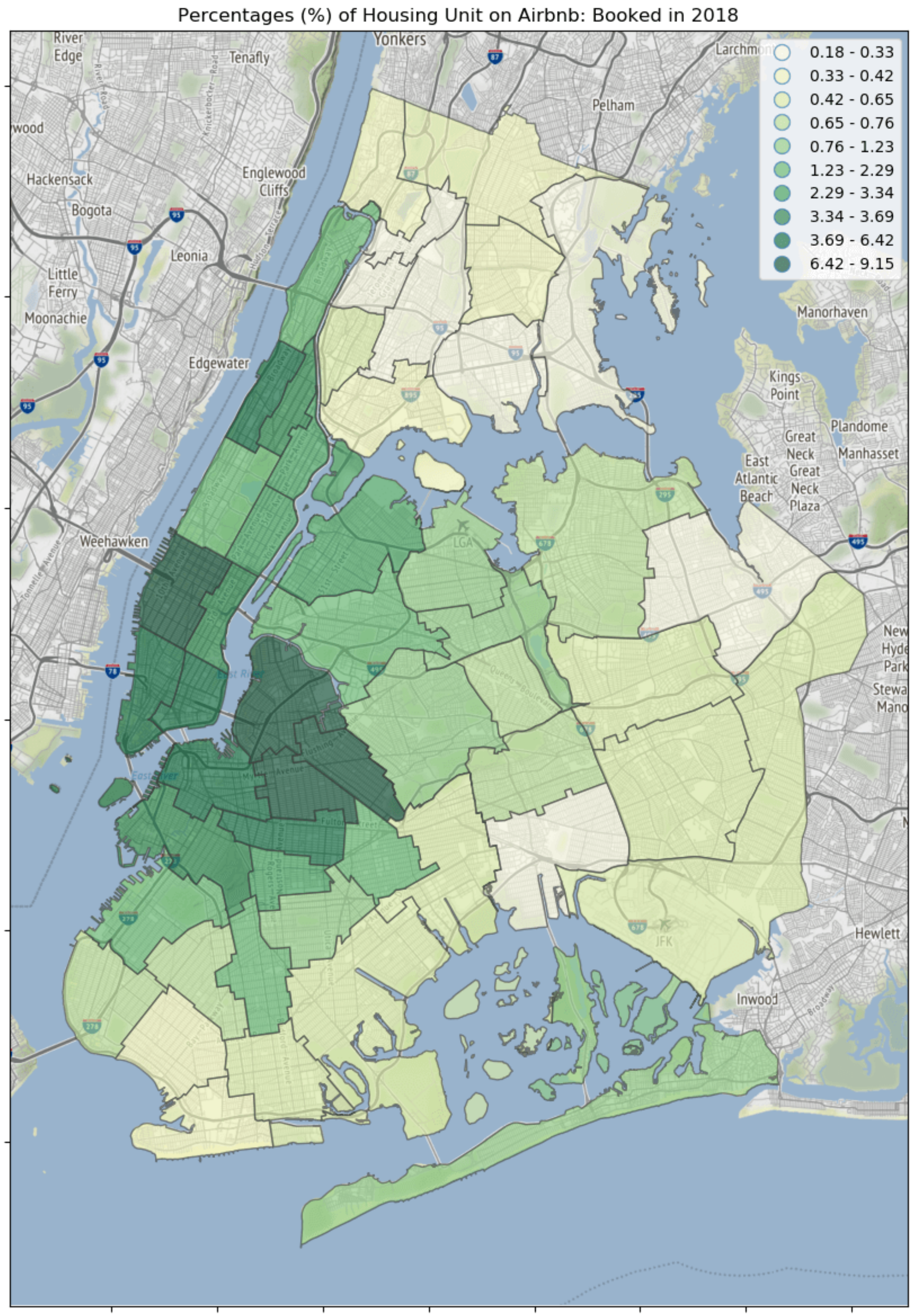


Figure 3: Percentage (%) of Housing Units on Airbnb
(Having At Least One Reservation in 2018)

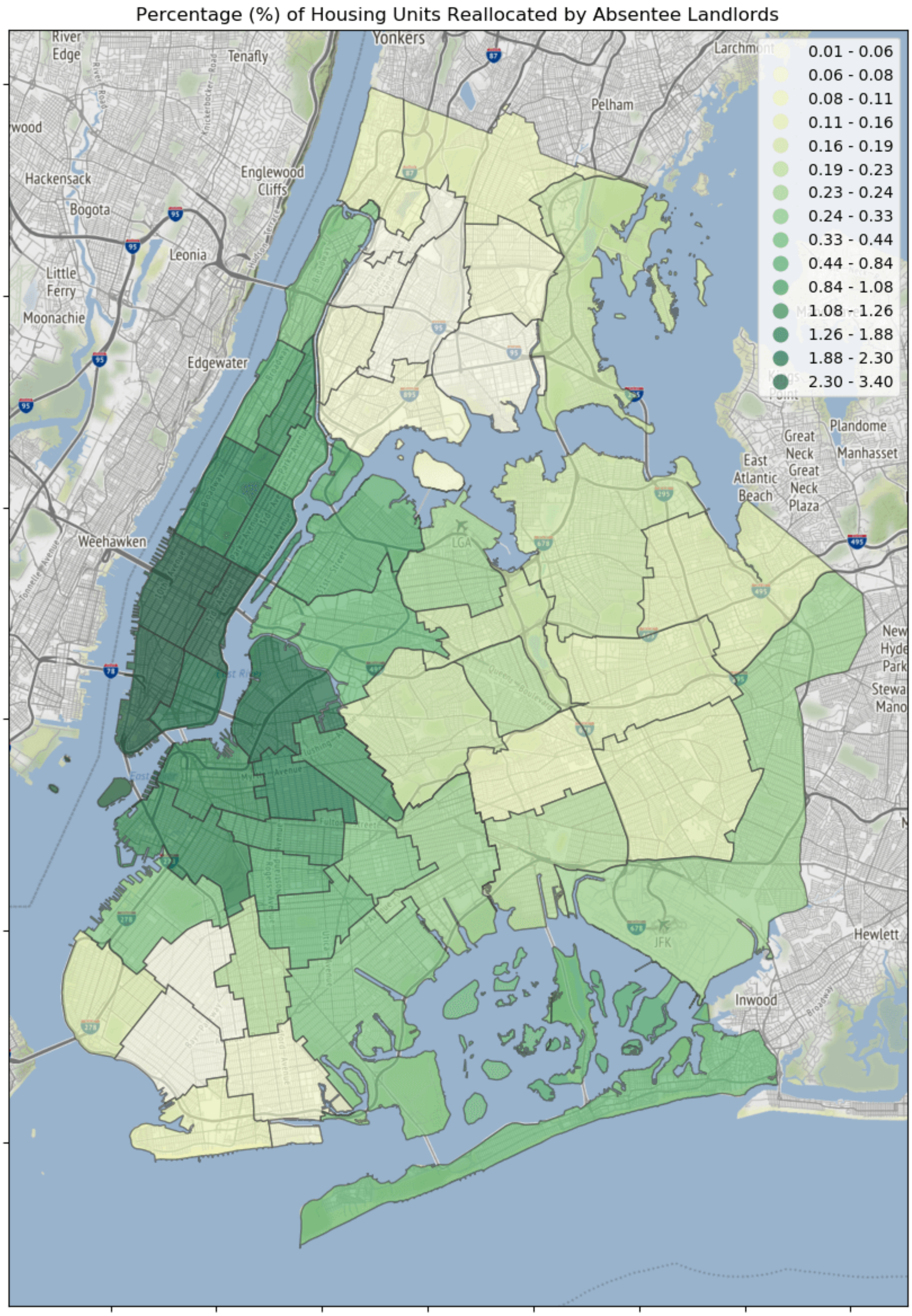
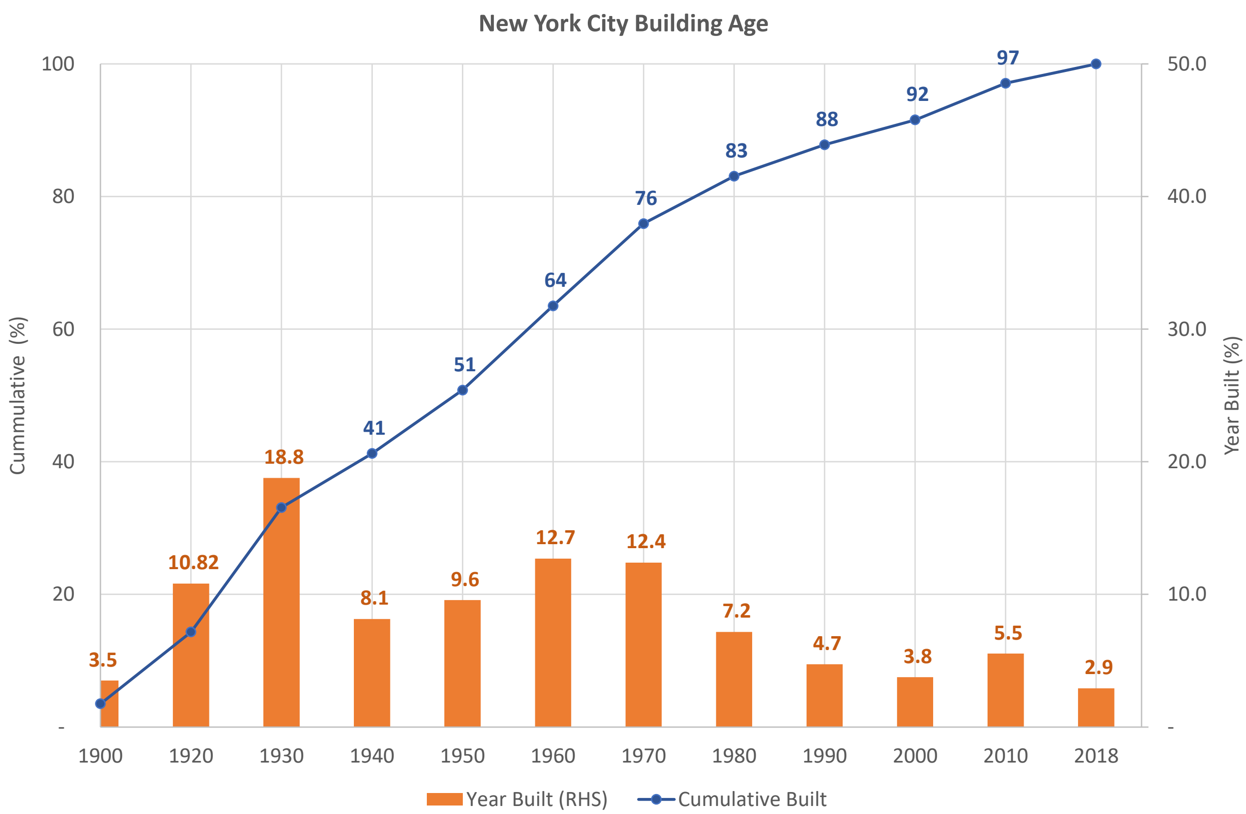


Figure 4: Percentage (%) of Rental Housing Units Reallocated
(Available on Airbnb for 180+ Days in 2018)

Figure 5: Building Age of NYC Housing Units

Data based on ACS 2017 and New York City Housing and Vacancy Survey (2017). Over 40% of the housing units were built prior to 1940. Housing construction since the 1980s has remained depressed. 88% of the housing were built prior to 1990. Only 2.9% of the housing stock was built post-2010, whereas 3.5% of the units were built prior to 1900.



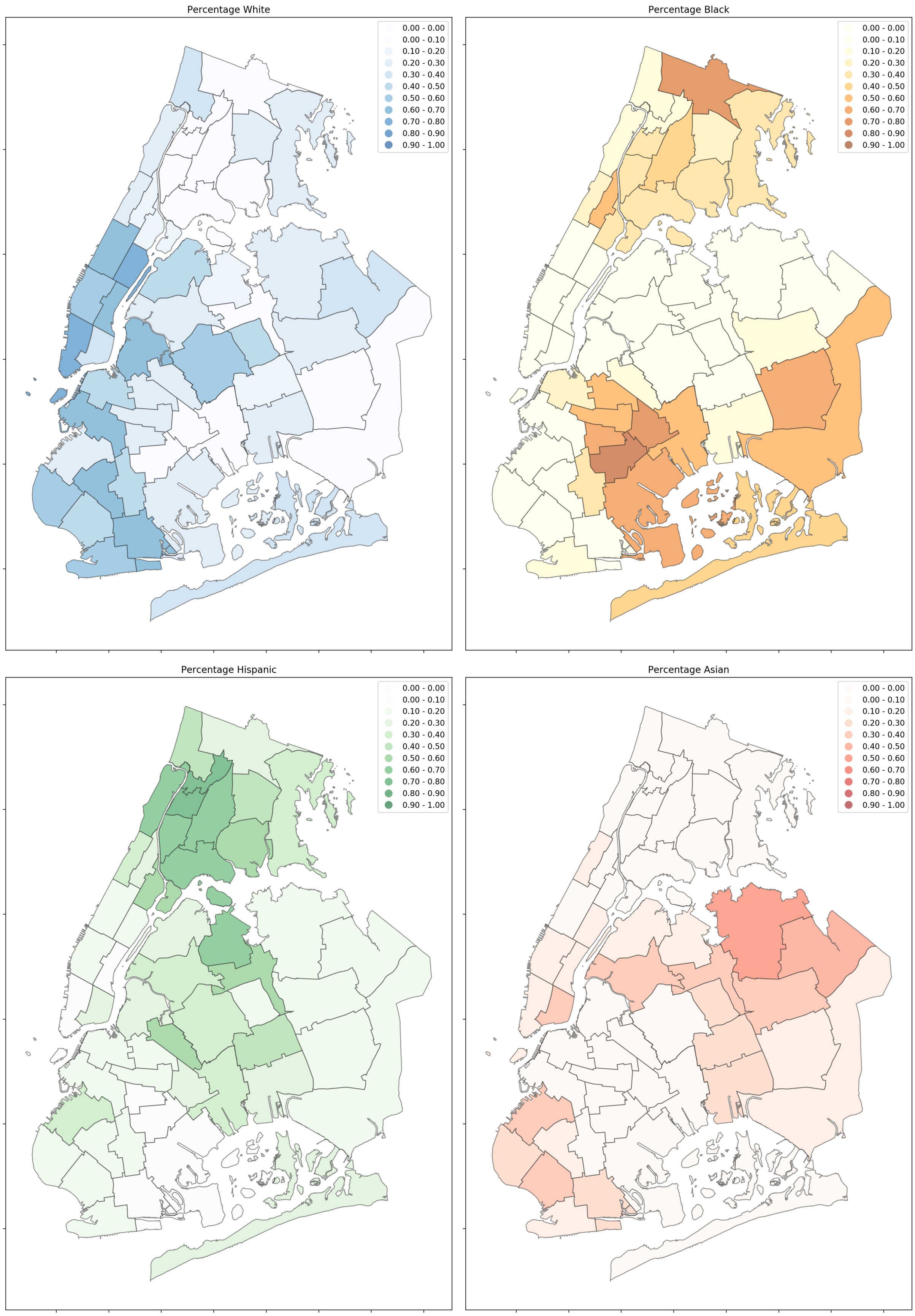


Figure 6: Race and ethnicity across neighborhoods

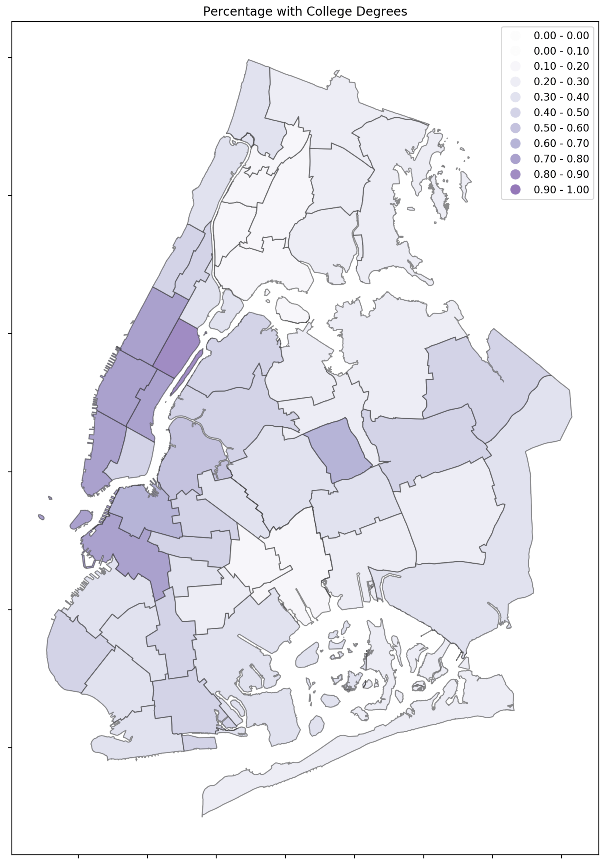


Figure 7: Education attainment across neighborhoods

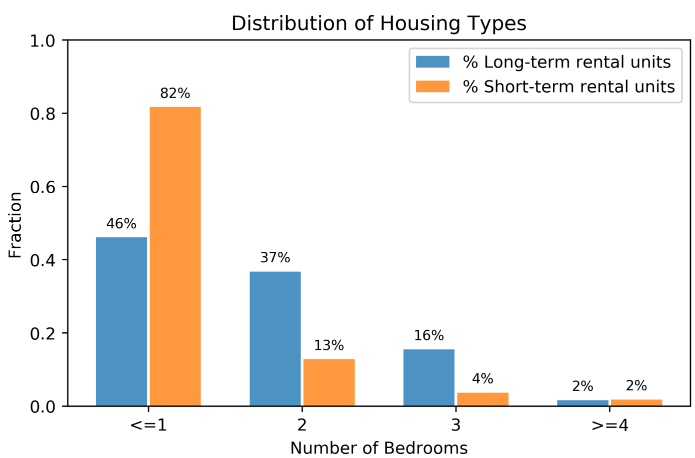
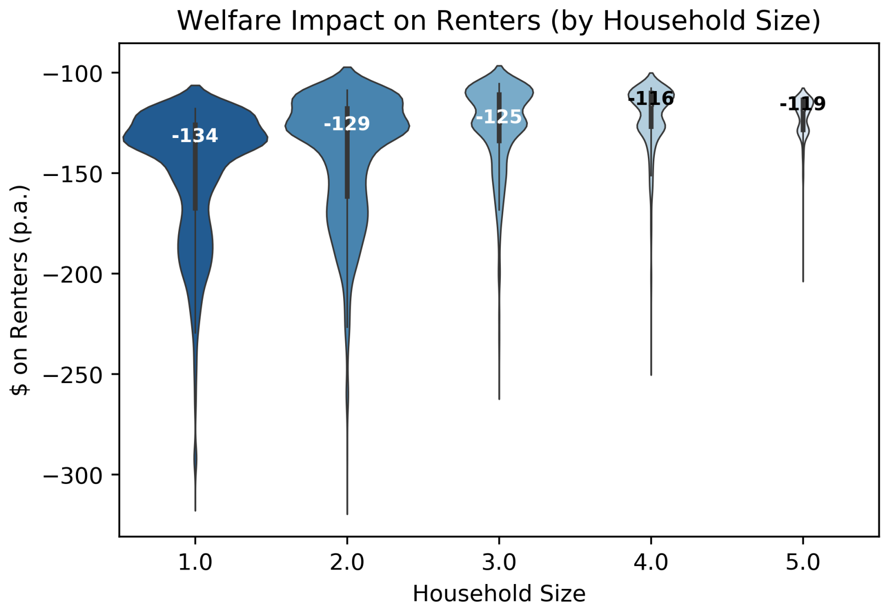
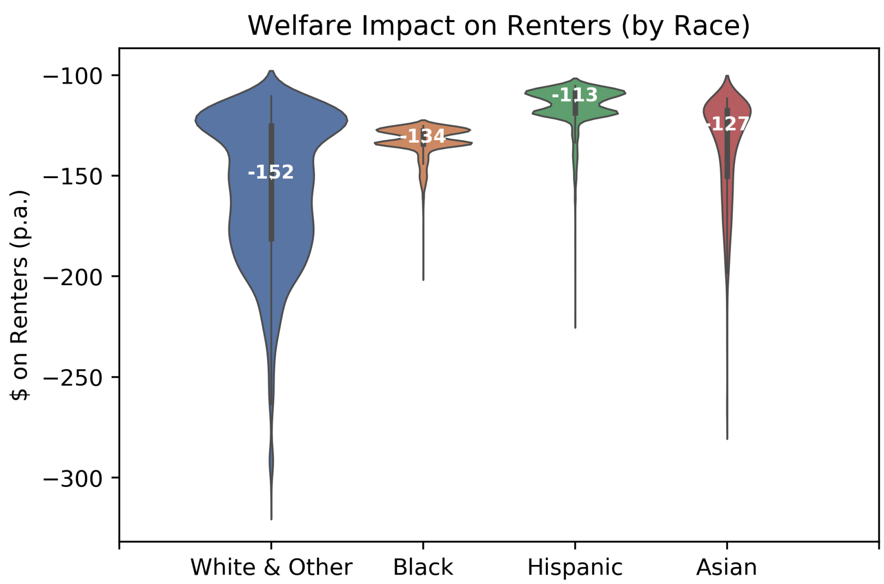


Figure 8: A comparison of housing types available on the long-term and the short-term rental market respectively. Notably, smaller housing types are much more prevalent on the short-term rental market than on the long-term rental market.

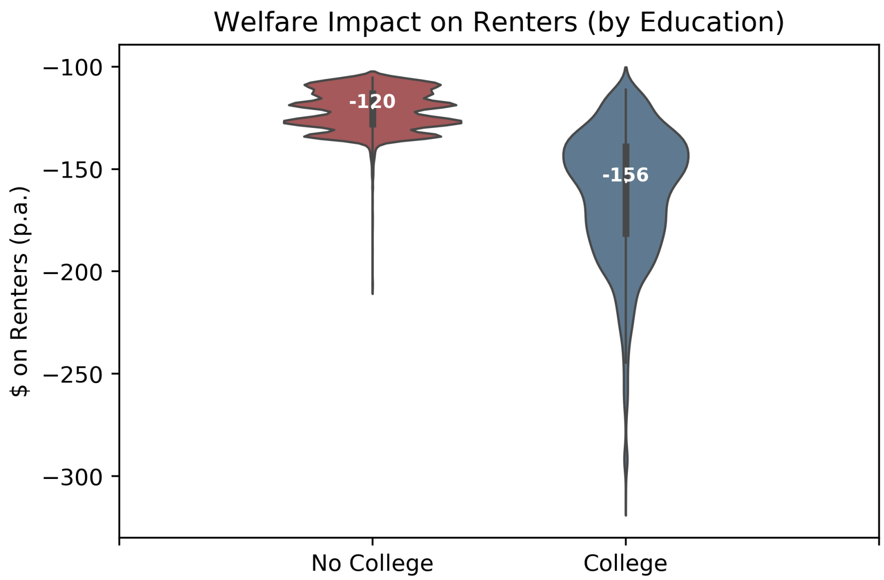


(a) Welfare impact via the rent channel by household size

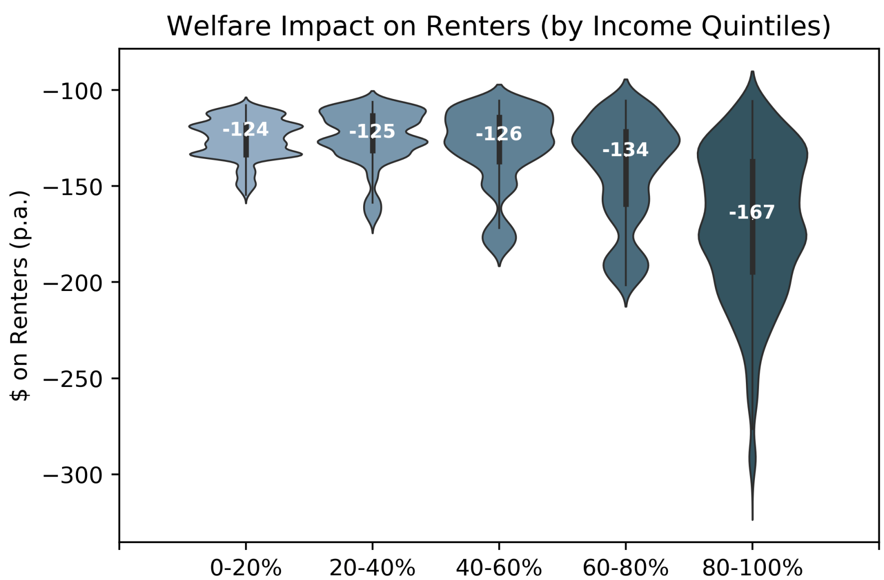


(b) Welfare impact via the rent channel by race and ethnicity

Figure 9: The welfare impact of Airbnb on renters *via the rent channel* by various demographic characteristics. The welfare loss is suffered more heavily by smaller households. In terms of race and ethnicity, the welfare loss is suffered more heavily by white renters. The labeled numbers indicate the category median. The width of each kernel density plot corresponds to the frequency of the category in the population. The mini-box plot in the center indicates inter-quartile range using the thick black line, and 1.5x inter-quartile range using the thin black line.



(a) Welfare impact via the rent channel by education



(b) Welfare impact via the rent channel by household income

Figure 10: The welfare impact of Airbnb on renters *via the rent channel* by various demographic characteristics. The welfare loss is suffered more heavily by educated and higher-income renters. The labeled numbers indicate the category median. The width of each kernel density plot corresponds to the frequency of the category in the population. The mini-box plot in the center indicates inter-quartile range using the thick black line, and 1.5x inter-quartile range using the thin black line.

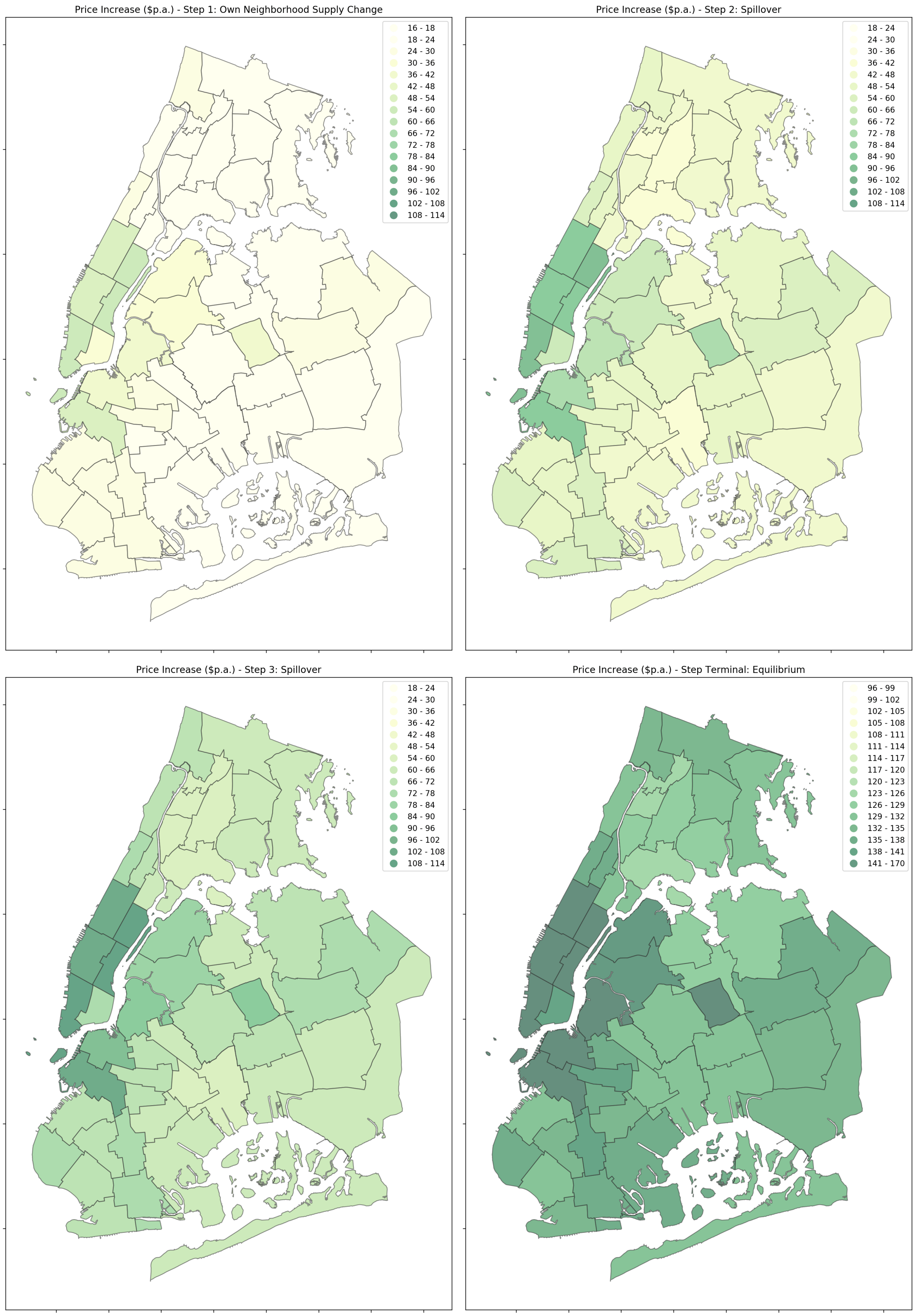


Figure 11: The equilibrium effects of supply restrictions (using successive steps of best responses to illustrate the equilibrating process)

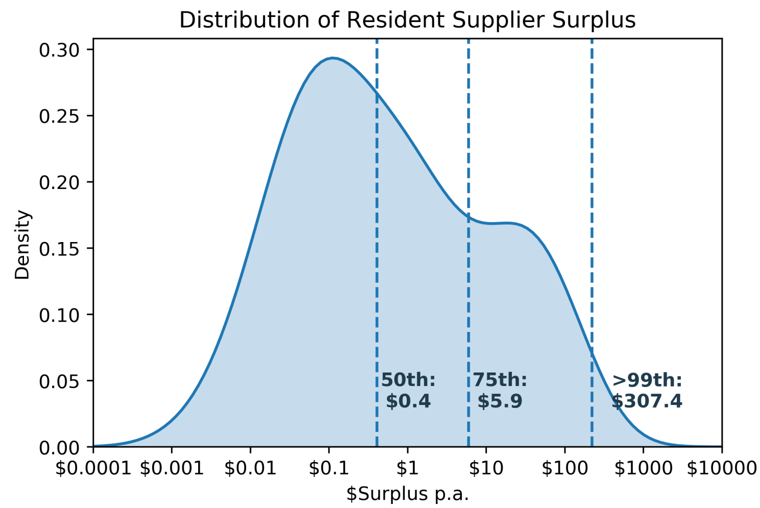


Figure 12: The kernel density plot of the supplier surplus through Airbnb, namely the welfare impact on renters via the *host* channel. Note that this plot is over the logarithm of the surplus. For the median household, the surplus from being an Airbnb supplier is immaterial at only \$0.4 p.a., reflecting the fact that most households do not participate in the sharing platform. The 70th percentile surplus is at \$5.9 p.a. The 90th percentile surplus is at \$45.6 p.a. However, in the very right tail (>99%), the surplus is substantial at \$307 p.a.

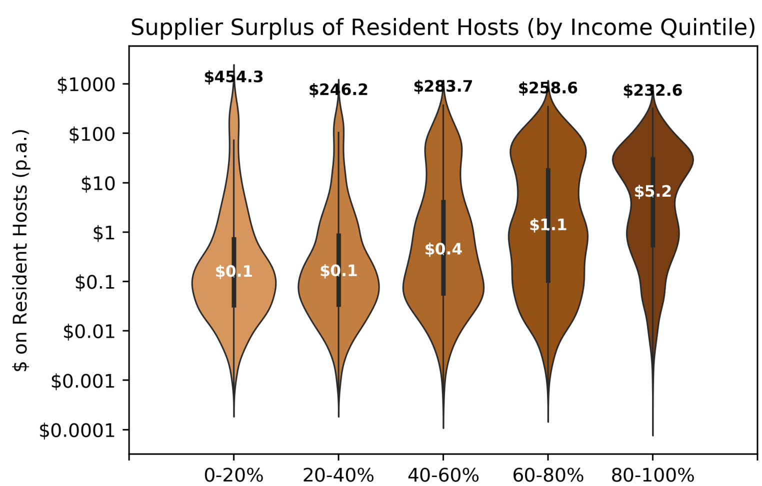


Figure 13: The welfare impact of Airbnb on renters via the *host* channel by income quintiles. Note that this plot is over the logarithm of the surplus on the vertical axis. The numbers in the middle of the density plot represent the category median, whereas the numbers at the top represent the conditional average above the 99th percentile. The median surplus from home-sharing is immaterial across all income levels, but higher-income groups still have a higher mean. In the tail, the lowest-income quintile accrue larger benefits.

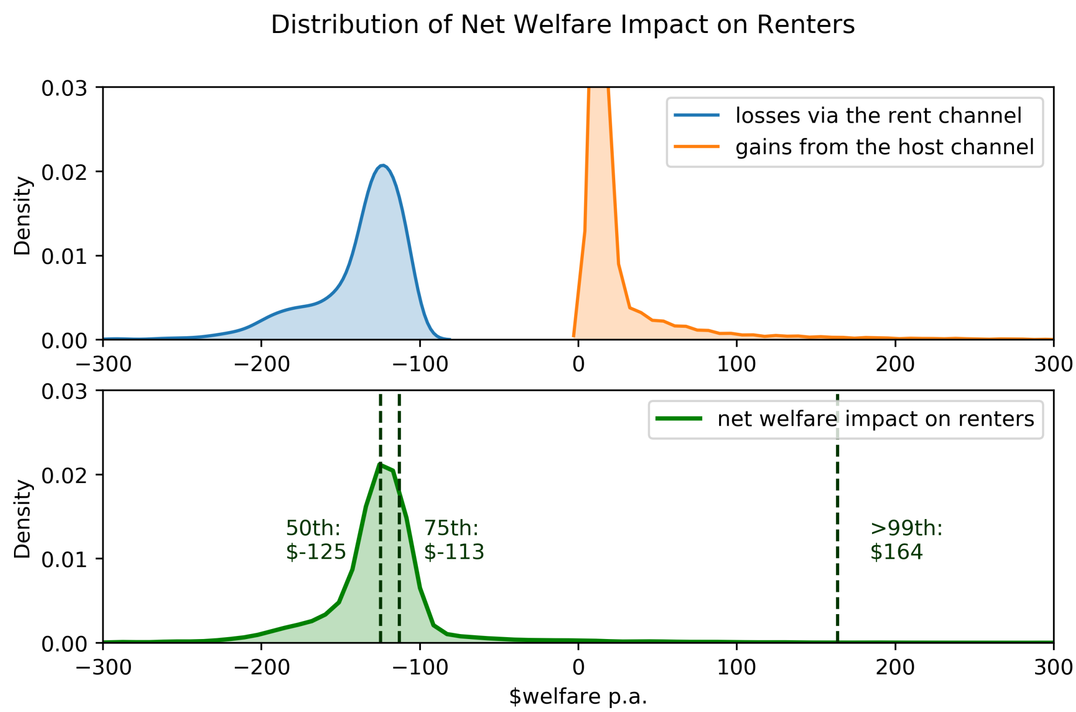


Figure 14: The net welfare impact of Airbnb on renters combining the loss from the rent channel and the gains from the host channel. The impact on the median renter is -$125 p.a. and the impact on the 75th-percentile renter is -$113 p.a. Note that there is a long tail on the right that the plot does not accommodate.

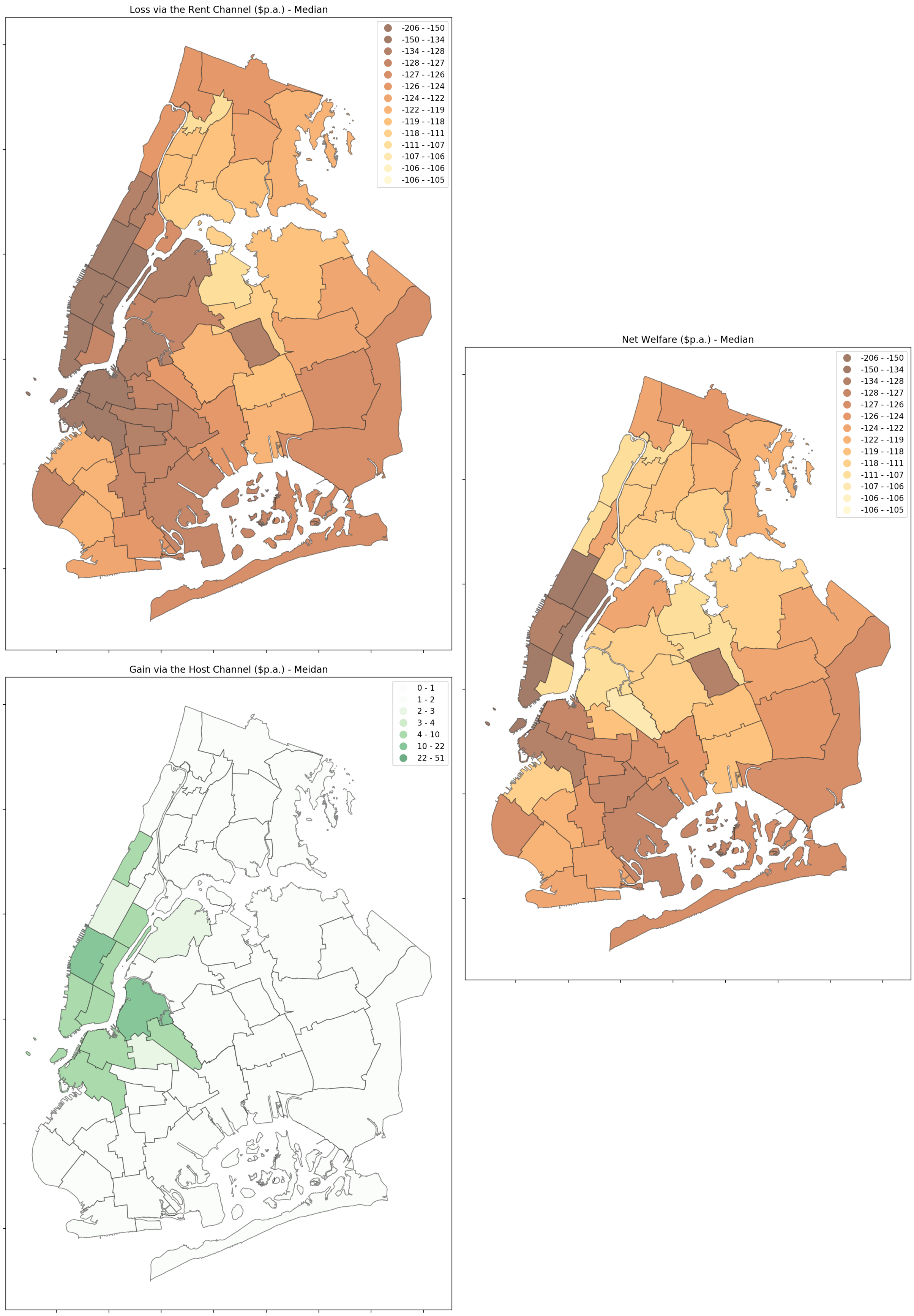


Figure 15: Net Welfare Impact by Neighborhood (Median): The median renter in all neighborhoods experiences a loss due to Airbnb as the rent channel dominates.

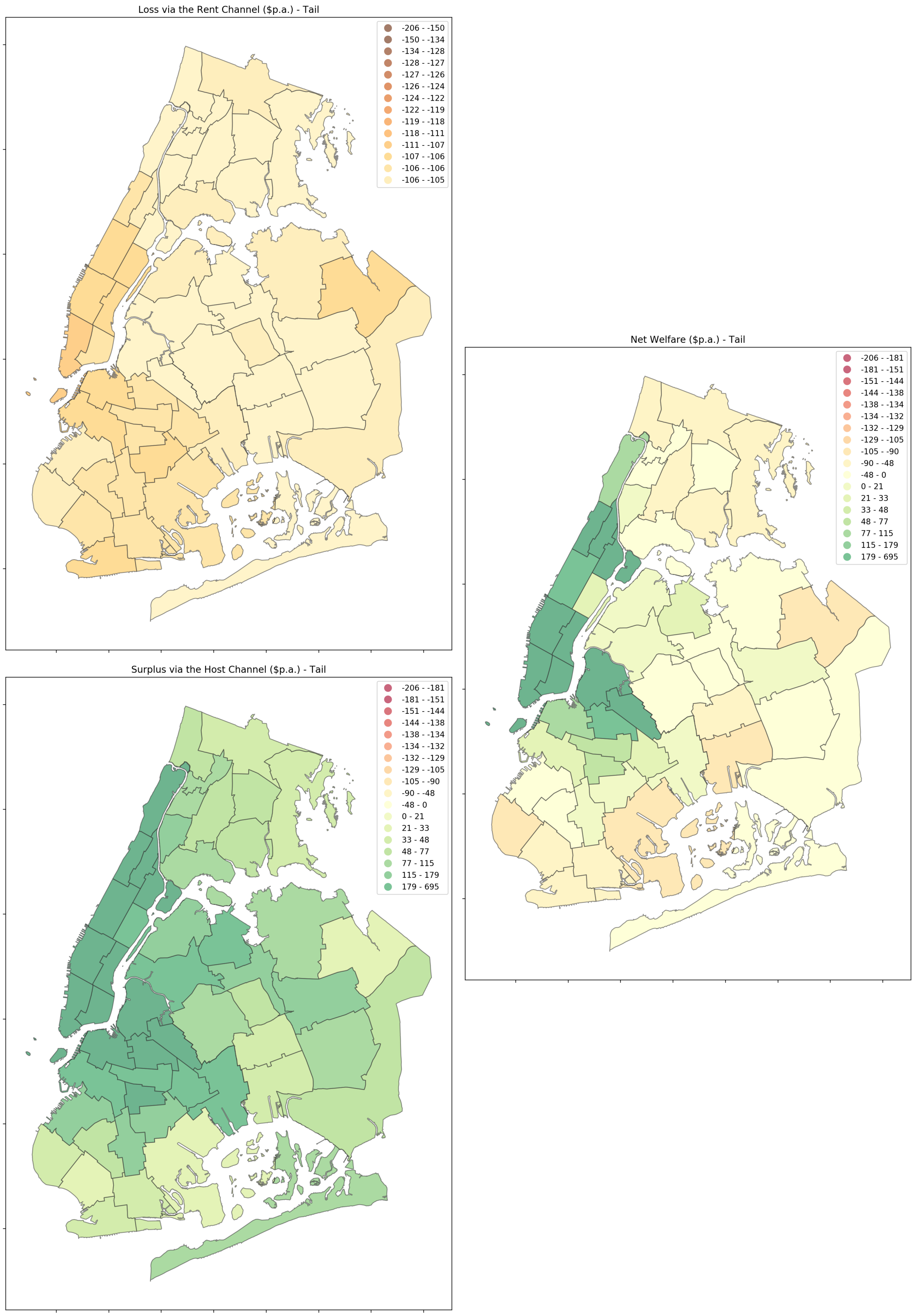


Figure 16: Net Welfare Impact by Neighborhood (Right Tail): For a small fraction of the renters, the gains from home-sharing can easily make up for the rent losses, especially in neighborhoods with high short-term rental demand.

8 Tables

Table 1: Breakdown of Airbnb Activity in NYC over 2018

Notes: This table summarizes the transactions on Airbnb in NYC over 2018. On Airbnb, a property is listed as one of three types: entire home/apt, private room, or shared room. Over 95% of the properties are listed as either "entire home/apt" or "private room", which are the focus of this paper. Overall, there are over 74,963 properties on Airbnb that have experienced at least one reservation, accounting for 2.2% of the housing units in the city.

Overview of Airbnb Activity in NYC in 2018

Listing Type	Bedroom(s)	Num Days Reserved (000s)	Num Properties (000s)	Average Daily Rate
Entire home/apt	All	3,204	38	$224
Entire home/apt	1	1,968	25	$178
Entire home/apt	2	838	9	$257
Entire home/apt	3	289	3	$350
Entire home/apt	4	82	1	$530
Private room	-	2,526	34	$86
Shared room	-	128	2	$59
Total		5,858	75	$156

Table 2: Income and Demographic Characteristics of NYC Renters

Notes: Based on the American Community Survey 2017 1-Yr estimates, the table summarizes the income and demographic characteristics of New York City renters across all boroughs, excluding Staten Island. Note that there are substantial variations in household income, education, race and ethnicity across the boroughs.

	All	**The Bronx**	**Brooklyn**	**Manhattan**	**Queens**
Annual Household Income ($k)					
0-20%	8	6	7	10	12
20-40%	24	17	22	31	32
40-60%	47	31	44	67	52
60-80%	83	54	78	126	83
80-100%	164	99	156	265	142
Education					
With College	38%	18%	36%	58%	32%
Race / Ethnicity					
White (non-Hispanic)	32%	7%	35%	47%	28%
African American	27%	36%	35%	16%	17%
Hispanic	32%	58%	21%	25%	31%
Asian	11%	2%	8%	12%	22%

Table 3: Reduced-Form Correlations between Airbnb and Neighborhood Characteristics

Notes: The *dependent variable* is the log of Airbnb share, namely, the share of housing units reallocated away from the long-term rental market. Observations are at the PUMA level. The neighborhood demographics are based on 2017 ACS data, whereas the neighborhood demographic changes are based on the changes from 2010 ACS to 2017 ACS. The correlation table shows that there are more Airbnb reallocation in neighborhoods that are relatively more white, educated, and higher-income. There is also some evidence that there are more Airbnbs in "gentrifying neighborhoods", as measured by improvements in the level of education.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Pct White	2.834***							
	(0.780)							
Pct Black		-0.336						
		(0.822)						
Pct Hispanic			-3.438***					
			(0.852)					
Pct Asian				-0.211				
				(1.493)				
Pct College					5.244***			
					(0.784)			
ln(Median Income)						1.803***		
						(0.382)		
Chg in Pct College							0.0974**	
							(0.0475)	
Chg in Pct Black								-0.153***
								(0.0513)
N	52	52	52	52	52	52	52	52

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Parameter Estimates for the Long-Term Rental Demand Model (Part I)

Notes: This table provides the heterogeneous coefficients estimated from *Step 1* of the long-term rental demand model. The coefficients on housing attributes are presented as willingness-to-pay in terms of monthly dollars. The coefficient on price is applied to the monthly rent ($k). The omitted categories are studios, those built prior to 1940, and buildings with fewer than 5 units. Neighborhood characteristics are standardized to variance 1. The standard errors for the ratios are computed using the delta method. I highlight the most significant demographic characteristics for each housing and neighborhood attribute. Notice strong demographic clustering along race, ethnicity, and education.

WTP ($)	Ln Income	HH Size	Black	Hispanic	Asian	College
Housing Characteristics						
One Bedroom	75.8	**255.1**	-40.3	-67.0	-177.1	-86.4
	(31.2)	**(78.6)**	(55.1)	(54.5)	(75.6)	(48.3)
Two Bedroom	59.1	**520.5**	98.1	-24.2	-273.4	-212.6
	(28.8)	**(156.4)**	(63.8)	(54.1)	(101.0)	(76.9)
Three Bedroom	32.4	**717.6**	143.8	-37.2	-329.1	-214.6
	(28.7)	**(214.9)**	(80.7)	(64.8)	(121.0)	(82.6)
Four Bedroom	85.0	**884.9**	-206.5	-328.3	-297.9	-244.9
	(66.7)	**(266.0)**	(172.3)	(171.9)	(170.2)	(134.0)
Built After 1980	22.4	-35.9	157.9	55.1	42.3	-27.8
	(14.7)	(13.7)	(56.4)	(33.3)	(36.7)	(25.0)
Built 1940-1980	-102.6	6.1	137.0	125.4	-84.8	58.6
	(34.3)	(10.1)	(55.6)	(50.6)	(47.5)	(32.6)
5+ Units	9.7	**58.6**	-3.8	-118.4	-110.6	-6.7
	(9.5)	**(18.6)**	(25.9)	(41.3)	(40.1)	(17.6)
Nbhd. Characteristics						
Pct Black (Std)	56.7	-47.9	**774.4**	**330.6**	272.1	67.6
	(20.3)	(15.9)	**(232.6)**	**(101.2)**	(86.6)	(27.9)
Pct Hispanic (Std)	56.3	-22.8	**376.9**	**469.3**	221.5	94.8
	(19.9)	(9.4)	**(115.8)**	**(141.5)**	(71.5)	(33.7)
Pct Asian (Std)	47.8	-14.3	98.2	138.0	**410.0**	-37.7
	(16.9)	(7.1)	(39.0)	(44.9)	**(123.9)**	(19.2)
Pct College (Std)	**145.9**	-54.0	185.9	37.0	93.8	**260.2**
	(45.7)	(18.5)	(68.1)	(32.2)	(44.7)	**(81.7)**
Inside NYC	**-337.8**	**-421.2**	120.0	29.1	299.0	-2.3
	(106.9)	**(128.6)**	(97.6)	(83.9)	(129.4)	(68.6)
Commuting Time (Std)	38.7	-6.3	127.7	50.7	210.4	8.4
	(19.9)	(11.9)	(56.9)	(43.0)	(76.8)	(33.1)
Monthly Rent	**0.33**	-0.03	-0.36	-0.23	-0.18	0.21
	0.10	0.02	0.13	0.10	0.10	0.08

Table 5: Parameter Estimates for the Long-Term Rental Demand Model (Part II)

Notes: This table provides the linear coefficients estimated from *Step 2* of the long-term rental demand model. The *dependent variable* is the mean utility of each housing type δ_h^L estimated from *Step 1* using individual-level data. If the long-term rental prices were assumed to be exogenous, column (1) shows that the price coefficient is biased. Column (2) shows the IV regression where I use the alternative equilibrium price vector at $\xi^L = 0$ as the instrument for the observed prices. In Column (3), I show the willingness-to-pay for each of the housing and neighborhood attributes for the average household.

	(1) OLS	(2) Instrumented	(3) ($) WTP Mo.
Monthly Rent ($k)	0.0213	-2.044***	
	(0.0341)	(0.609)	
One-Bedroom	0.425***	0.929***	454.5***
	(0.0447)	(0.188)	(78.2)
Two-Bedroom	0.528***	1.325***	648.2***
	(0.0465)	(0.280)	(93.5)
Three-Bedroom	0.271***	1.392***	681.0***
	(0.0555)	(0.393)	(76.7)
Four-Bedroom	-0.179***	0.904*	442.3*
	(0.0668)	(0.505)	(162)
Built After 1980	-0.114***	0.139	68.2
	(0.0402)	(0.145)	(60.9)
Built 1940 to 80	-0.00917	-0.242**	-118.4**
	(0.0337)	(0.105)	(43.9)
5+ Units	0.00182	-0.209**	-102.3**
	(0.0282)	(0.0974)	(41.2)
Commuting Time (Std)	0.119***	-0.782***	-382.6***
	(0.0215)	(0.279)	(28.2)
Inside NYC	-1.026***	2.536**	1240.7***
	(0.0683)	(1.036)	(147)
N	1050	1050	1050

Table 6: Parameter Estimates for the Short-Term Rental Supply Model

Notes: This table provides the estimated parameters for the short-term rental supply by resident hosts, using the MPEC procedure. The standard errors are clustered at the neighborhood level. In columns (1) and (2), prices are assumed to be exogenous and not instrumented. In columns (3) and (4), the total hotel bookings in the city (lagged by seven years) are used as the instrument for prices in the short-term rental market. The instrument is strong, with an F-stat of 25.4 in column (4) when controlling for month FE, day of week FE and holiday FE (including Christmas and New Year's Eve) in the corresponding linear specification. Column (5) provides the costs in dollar terms (per diem). Overall, the average cost to host is high, and the low-cost suppliers are those with a college degree, young, and have no children. The average price elasticity is 5.96. Since the interaction between household income and price is negative, low-income households are more elastic, averaging 6.70 for one-standard deviation lower in income.

		(1) Naïve	(2) Naïve	(3) IV	(4) IV	(5) ($) per diem
Linear Coef.	*Non-Linear Coef.*					
Price		0.006	0.007	0.052	0.056	
		(0.002)	(0.001)	(0.002)	(0.002)	
	x ln(income)	-0.018	-0.018	-0.018	-0.011	
		(0.001)	(0.002)	(0.003)	(0.006)	
Cost		15.44	15.51	22.07	21.36	224.3
		(0.10)	(0.09)	(0.12)	(0.11)	(12.7)
	x Has College	-1.17	-2.55	-3.47	-3.27	-58.9
		(0.68)	(0.24)	(0.27)	(0.25)	(4.8)
	x Has Children	2.40	2.58	1.95	2.60	46.7
		(0.42)	(0.36)	(0.53)	(0.44)	(8.1)
	x Age (yr)	0.094	0.093	0.091	0.097	1.8
		(0.005)	(0.005)	(0.006)	(0.006)	(0.1)
	x ln(income)	0.24	-0.14	-0.39	-0.29	-5.1
		(0.09)	(0.13)	(0.26)	(0.48)	(8.7)
Quad. Time		Yes	Yes	Yes	Yes	
Month FE		No	Yes	No	Yes	
Day of Week FE		No	Yes	No	Yes	
Holiday FE		No	Yes	No	Yes	
N		75,895	75,895	75,895	75,895	

Table 7: Counterfactual using Actual Airbnb Penetration vs. Uniform Airbnb Penetration

Notes: This table compares the counterfactual analysis using the actual Airbnb listing data with an alternative counterfactual analysis using a hypothetical scenario where the penetration of Airbnb is uniform across space and housing types while holding the total supply of Airbnb listings the same. This comparison illustrates that the primary driver of the welfare differences is due to the geography of actual Airbnb listings. The remaining difference in the counterfactual welfare with uniform Airbnb activity captures the higher willingness-to-pay for housing attributes by higher-income and more educated households.

	Actual Penetration		**Uniform Penetration**	
($ p.a.)	Median	Difference	Median	Difference
Household Size				
Household Size = 1	(134)		(142)	
Household Size = 5	(119)		(141)	
Small vs. Large Households		**(15)**		**(1)**
Race / Ethnicity				
White	(152)		(137)	
Black	(134)		(150)	
Hispanic	(113)		(130)	
Asian	(127)		(132)	
White vs. Hispanic		**(39)**		**(7)**
Education				
With College	(156)		(141)	
Without College	(120)		(132)	
With vs. Without College		**(36)**		**(9)**
Household Income				
Highest Quintile	(167)		(144)	
Lowest Quintile	(124)		(135)	
Highest vs. Lowest Income Quintile		**(43)**		**(9)**

Table 8: Estimated Own- and Cross- Price Semi-Elasticities

Notes: Entry (i, j) corresponds to the percentage (%) of share changes in neighborhood j when the price of housing in neighborhood i increases by $1,000 per month. The table here shows only a subset of the neighborhoods for illustration, whereas the full matrix is 52 x 52. For example, this table shows that when price increases in Chelsea, Clinton & Midtown, households there will substitute away towards similar neighborhoods such as the Upper West Side. However, the substitution towards Forest Hill (an educated and predominantly white neighborhood) is much larger Jackson Heights (a less-educated and predominantly Hispanic neighborhood), even though they are geographically closely located in Queens.

	The Bronx Hunts Point, Longwood & Melrose	*Manhattan* Central Harlem	*Manhattan* Upper West Side & West Side	*Manhattan* Chelsea, Clinton & Midtown	*Queens* Jackson Heights & North Corona	*Queens* Forest Hills & Rego Park
Hunts Point, Longwood & Melrose	(21.9)	0.55	0.09	0.15	0.32	0.20
Central Harlem	0.58	(23.7)	0.20	0.32	0.18	0.30
Upper West Side & West Side	0.10	0.20	(21.0)	1.02	0.10	0.69
Chelsea, Clinton & Midtown	0.12	0.21	0.70	(21.6)	0.12	0.71
Jackson Heights & North Corona	0.62	0.33	0.19	0.32	(18.8)	0.35
Forest Hills & Rego Park	0.15	0.21	0.54	0.81	0.14	(37.5)

Table 9: Net Welfare Impact by Household Income

Notes: This table compares the welfare impact of Airbnb on renters by household income quintiles. Overall, higher-income households experience larger losses on average and experience smaller gains in the tail.

Welfare Impact by Renter Income Quintile ($ p.a.)

Loss via the Rent Channel	Mean	Median	P25	P75	>P99
0-20%	-125	-124	-134	-118	-115
20-40%	-124	-125	-131	-113	-106
40-60%	-130	-126	-137	-114	-106
60-80%	-142	-134	-159	-122	-105
80-100%	-169	-167	-195	-137	-106

Gain via the Host Channel	Mean	Median	P25	P75	>P99
0-20%	12	0.1	0.0	0.7	454
20-40%	7	0.1	0.0	0.8	246
40-60%	14	0.4	0.1	4.0	284
60-80%	20	1.1	0.1	17.9	259
80-100%	24	5.2	0.5	30.0	233

The Net Welfare Impact	Mean	Median	P25	P75	>P99
0-20%	-114	-122	-131	-114	319
20-40%	-117	-118	-128	-111	101
40-60%	-116	-119	-130	-109	137
60-80%	-122	-125	-139	-112	109
80-100%	-146	-144	-174	-122	73

Table 10: Net Welfare Impact by Household Demographics

Notes: This table compares the median welfare impact of Airbnb on renters by household characteristics. Overall, educated and white renters experience greater losses.

	Median Impact $p.a.			Tail Impact $p.a.		
	Loss via Rent	Gain via Host	Net Impact	Loss via Rent	Gain via Host	Net Impact
Overall	-128	0.4	**-125**	-109	307	164
Education						
Without College	-120	0.1	**-120**	-105	16	-98
With College	-156	10.8	**-136**	-112	393	253
Race						
Asian	-127	0.7	**-119**	-112	381	245
Black	-134	0.2	**-129**	-125	227	85
Hispanic	-113	0.1	**-111**	-105	232	107
White & Other	-152	1.9	**-130**	-111	326	179

Table 11: Welfare Impact on All Participants

Notes: This table summarizes the aggregate welfare impact for all relevant participants of the long-term and short-term rental market. The black numbers represent outputs from the model. The gray numbers represent back-of-envelope approximations. In particular, absentee landlords' gains from hosting is based on the difference between their revenue from Airbnb less the forgone rent, assuming the same profitability as hotels. The net welfare impact on tourists and hotels is based on Farronato and Fradkin (2018), which estimates an average consumer surplus of $42 per room night and a decline of 5% in hotel profitability. Hotel usage and profitability are based on data from Smith Travel Research. Although the net welfare impact on renters is negative due to the transfer to property owners, the net impact inclusive of owner-occupiers and landlords is positive.

Welfare Impact on Various Stakeholders ($mm p.a.)			
Renters' loss to absentee landlords	-200		
Renters' loss from displacement	-1		
Renters' gain from hosting	23		
Net Impact on Renters	-178		
Owner occupiers' gain from hosting		5	
Absentee landlords' gain from renters		200	
Absentee landlords' gain from hosting		46	
Net Impact on Housing Market Participants		73	
Net welfare on tourists and hotels			126
Net Impact			198

A Appendix Figures

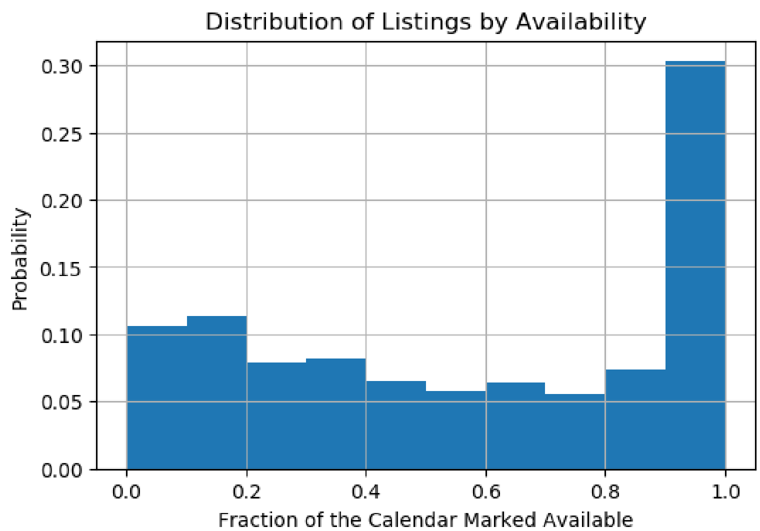


Figure A.17: Distribution of Listings by Calendar Availability

Figure A.18: The daily number of reservations of private rooms sold on Airbnb in New York City.

Figure A.19: The daily average price of private rooms sold on Airbnb in New York City.

B More on LTR Demand Estimation

The main idea is that one can use a supply-side pricing equation to produce a price instrument. The key intuition is that the availability of similar housing characteristics in the market has an impact on the equilibrium price of a given house, but is uncorrelated with the unobserved quality. Moreover, if I have two or more housing characteristics, then their relative availability allows this identification strategy to work even with only one cross-section of the market.

In this section, I will first describe the model, followed by the identification strategy, and end with some findings from a simulation study.

B.1 Model

Consider the following model with N households indexed by i and M homes indexed by j in a given housing market. Each home is endowed with a vector of physical features $X_{j,k}$ with $k = 0, \ldots, K$, where there are at least two features $K \geq 2$. Let $X_{j,0}$ denote the indicator for the outside option, with its price normalized to zero. Moreover, each home has an unobserved quality component ξ_j.

There is a random coefficient $\nu_{i,k}$ associated with each of the K features, drawn randomly from a normal distribution $\sigma \times \mathcal{N}(0, I)$. Here, I assume σ is known to simplify the exposition.[79]

The utility for household i renting home j is as follows:

$$U_{i,j} = \alpha p_j + \sum_k \beta_{i,k} X_{j,k} + \xi_j + \epsilon_{i,j} \tag{B.1}$$

where $\beta_{i,k} = \beta_k + \nu_{i,k}$.

Given that each household maximizes its utility, the choice probability of a household i over home j becomes:

$$P_{i,j}(p, \mathbf{X}; \Theta) = \frac{exp(\delta_j + \lambda_{i,j'})}{\sum_{j'} exp(\delta'_j + \lambda_{i,j'})} \tag{B.2}$$

$$\delta_j(p, \mathbf{X}; \Theta) = \alpha p_j + \sum_k \beta_k X_{j,k} + \xi_j \tag{B.3}$$

$$\lambda_{i,j}(p, \mathbf{X}) = \sum_k \nu_{i,k} X_{j,k} \tag{B.4}$$

where $\Theta = (\alpha, \vec{\beta}_k, \xi)$.

The model is closed by a sorting equilibrium where the price p^e clears the market: The

[79] In the actual model, it is captured by the coefficients in front of household demographics, which are estimated offline first using individual level choice data.

demand for each home equals the observed supply, namely $\forall j : s_j^F = 1$.

$$\forall j : \sum_i P_{i,j}(p^e, \mathbf{X}; \Theta) = 1 \tag{B.5}$$

B.2 Estimation

The key identification assumption is that the unobserved quality ξ is *independent* of the physical features of the home $\mathbf{X}$.

As a result, the unobserved quality ξ is uncorrelated with the price instrument z, which is constructed as the market clearing price with ξ set to zero.

$$\forall k : \mathbb{E}[\xi X_k] = 0 \tag{B.6}$$

$$\mathbb{E}[\xi z] = 0 \tag{B.7}$$

$$\forall j : \sum_i P_{i,j}\big(z, \mathbf{X}; (\alpha, \vec{\beta}, \xi = 0)\big) = 1 \tag{B.8}$$

$$\forall j : \sum_i P_{i,j}\big(p^e, \mathbf{X}; (\alpha, \vec{\beta}, \xi)\big) = 1 \tag{B.9}$$

Remark B.1. Why does the instrument work? The key intuition is that the supply-side model, namely that the demand clears the fixed supply of homes, implies that homes with rarer characteristics are going to have higher equilibrium prices, *compared with* homes with housing characteristics that are more common. And this component of variation (the scarcity premium) is uncorrelated with the unobserved quality of the particular home.

Suppose that there are only two relevant housing characteristics in the market: X_1 indicates that a home was built within the last 5 years, which is relatively rare in New York City, and X_2 indicates that a home has two bedrooms (as opposed to just one), which is more common. Both are desirable features. Without loss of generality, assume that the mean utilities for the two features are the same $\beta_1 = \beta_2$.

The key identification assumption is that the price premium observed for homes in brand-new buildings is higher than the price premium for homes with an extra bedroom; this is not because homes with two bedrooms have unobservably lower quality but rather because it is a much more common housing characteristic than being brand-new.

Remark B.2. It is essential that there be a distribution of preferences over the physical features of the homes. This is because for the IV strategy to work, the model has to produce equilibrium prices that are higher for rarer features, even if the valuations for it by the average household (namely, the household with $\nu_i = 0$) might be completely identical. In other words, the distribution of preferences over characteristics means that those who care a lot about a rare characteristic will bid up its price in equilibrium. In practice, as it is very likely that households of different income levels will have different price sensitivities, there will be a distribution over the willingness-to-pay $\beta_{i,k}/\alpha_i$ for housing attributes.

Remark B.3. It is also essential that there are at least two relevant housing characteristics $K \geq 2$ in the market for the estimation to work with just one cross-section. Otherwise, the concept of the rarity of a characteristic is undefined. Because I can compare the rarity of one housing characteristic with another housing characteristic in just one cross-section, it is also the reason why the instrument works even with just one cross-section of the market.

Remark B.4. It is also important that $\mathbf{X}$ is independent of ξ. Because the instrument z is a non-linear function of $\mathbf{X}$ as well as the parameters, the independence assumption ensures that z will be uncorrelated with ξ as a result.

Remark B.5. In general, there is a hump-shaped relationship between the mean utility (δ_j) and the price (or the price instrument): Namely, the mean utility for very expensive and very cheap homes tends to be lower than the mean utility of an average home. For expensive homes with rare and sought-after characteristics, prices will be too high compared to what an average household is willing to pay, resulting in low δ. On the other end, homes without these desirable characteristics are not going to have prices that are low enough for the average household to justify its lack of these characteristics. This is an inherent feature of the model, rather than evidence of the presence of ξ.

B.3 Simulation Study

In this section, I generate a very simple dataset with only two binary features X_1, X_2 with known parameters, where the first housing characteristic X_1 is much rarer than the second one X_2, even though the mean utilities on them are the same.

I show that the price instrument constructed in Eq (B.8) indeed recovers the true parameter, unlike the OLS. Moreover, I will also show that the typical hedonic regression will result in estimated WTP for amenities significantly different from the true model parameters, whereas the proposed IV estimation strategy does recover the WTP for the average household as specified by the model.

B.3.1 Simulation Set-up

More specifically, I have $N = M = 200$. Here is the breakdown of their characteristics with one home being the outside option (indicator by X_0):

(i) 20 homes have $X_1 = 1, X_2 = 1$

(ii) 20 homes have $X_1 = 1, X_2 = 0$

(iii) 120 homes have $X_1 = 0, X_2 = 1$

(iv) Remaining 39 homes have $X_1 = 0, X_2 = 0$

The true parameters of the simulation are set as follows:

$$\alpha = -1, \quad \beta_0 = 3, \quad \beta_1 = 1, \quad \beta_2 = 1 \tag{B.10}$$

Moreover, the unobserved quality ξ is drawn from a normal distribution with a standard deviation of 1/200. ν_i are drawn from a normal distribution with a standard deviation of 1 for both characteristics. With the model fully specified as above, I can solve the model by computing the equilibrium price $p^e(\mathbf{X}; \alpha, \vec{\beta}, \xi)$. Then, the mean utility δ is computed as follows:

$$\delta_j = \alpha p_j^e + \beta_0 X_{j,0} + \beta_1 X_{j,1} + \beta_2 X_{j,2} + \xi_j \tag{B.11}$$

B.3.2 Simulation Results

I compute the price instrument $z(\mathbf{X}; \alpha, \vec{\beta}, \xi = 0)$ that clears the market. I checked that the price instrument is indeed uncorrelated with the unobserved quality with $cov(z, \xi) = 0$.

I find that the IV recovers the true price coefficient, whereas the OLS produces a biased estimate of them, as shown in B.1.

I also find that the estimation strategy correctly recovers the true value of both amenities $(-\beta_k/\alpha)$, whereas a hedonic regression of price on amenities greatly overestimates the value of the rarer amenity X_1 and underestimates the value of the more common amenity X_2, as illustrated in Table B.2.

Note that the result *does not* require $\beta_1 = \beta_2$. I can vary the parameter vector to different values and the estimation strategy will still work, as long as there are at least two characteristics.

Table B.1: Regression on Mean Utility

The *Dependent variable* is the mean utility δ_j. The price instrument is constructed using the market clearing prices assuming $\xi = 0$. The F-stat of the first stage is well above 10 (at 306.0). The OLS produces a biased estimate of the price coefficient, whereas the IV recovers it.

	(1) OLS	(2) First Stage	(3) IV
Price	-0.635***		-1.042***
	(0.0363)		(0.0589)
Price Instrument		0.960***	
		(0.0549)	
X1	0.498***	0.0549	1.057***
	(0.0499)	(0.0754)	(0.0810)
X2	0.715***	0.0316	1.033***
	(0.0283)	(0.0428)	(0.0460)
Inside Option	1.907***	0.120	3.125***
	(0.109)	(0.164)	(0.176)
N	200	200	200

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.2: Estimation of the Willingness to Pay for Amenities

In this table, I compare the amenity value estimated by the model $(-\beta_k/\alpha)$ with the hedonic regression. In particular, even though the mean utility of the two characteristics X_1 and X_2 are identical, the relative scarcity of X_1 pushes the hedonics to produce a much greater coefficient.

	(1) Hedonic	(2) Sorting
X1	1.374***	1.015***
	(0.00139)	(0.0204)
X2	0.781***	0.991***
	(0.00122)	(0.0120)
Inside Option	2.994***	2.999***
	(0.00110)	(0.000760)
N	200	200

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

C More on STR Supply Estimation

In this section, I provide a number of relevant computational details about the estimation procedure used to estimate the supply system. Following Dubé, Fox and Su (2012), I formulate the GMM objective function as a mathematical program with equilibrium constraints (MPEC). With the problem appropriately defined, I provide the analytical Jacobian and Hessian used to estimate the parameters, where I highlight the sparsity features of the problem that make the computation reasonably efficient.

C.1 Problem Formulation

Without loss of generality, let q index a total of Q markets. Let $Z_q = [\,p_q^{IV}\ X_q^R\,]^T$ be a vector of exogenous shifters and $X_q = [\,p_q\ X_q^R\,]^T$ be the endogenous shifters. Let $Z = [Z_1, \ldots, Z_Q]^T$ and $X = [X_1, \ldots, X_Q]^T$ be the data matrix of size $Q \times (X+1)$. Let D_q denote the empirical distribution of the demographic characteristics in market q.

The requisite moment condition is

$$\mathbb{E}\big[(\delta_q - \theta_1^T X_q)^T Z_q\big] = 0 \tag{C.1}$$

To solve for the supply system coefficients, I formulate the problem as follows

$$\min_{(\delta,\theta_1,\theta_2,\eta)} \quad \eta^T W \eta \tag{C.2}$$

$$\text{s.t.} \quad \forall q:\ S_q(\delta_q, \theta_2; X_q, D_q) = S_q^o \tag{C.3}$$

$$\eta = Z'(\delta - \theta_1^T X) \tag{C.4}$$

As such, I denote the Lagrangian of the problem as follows

$$f_q(\delta_q, \theta_2; X_q, D_q) = S_q(\delta_q, \theta_2; X_q, D_q) - S_q^o \tag{C.5}$$

$$g(\delta, \theta_1; Z, X) = \eta - Z^T(\delta - \theta_1^T X) \tag{C.6}$$

$$G(\eta; W) = \eta^T W \eta \tag{C.7}$$

$$\mathcal{L}(\delta, \theta_1, \theta_2, \eta) = G(\eta; W) + \sum_q \lambda_f^q f_q(\delta, \theta_2; X_q) + \lambda_g^T g(\delta, \theta_1; Z, X) \tag{C.8}$$

Note that θ_1 denotes the linear coefficients of the model, whereas θ_2 denotes the non-linear coefficients of the model. Further, let $\theta_2 = [\pi_b^k]$ where $b = 1, \ldots, (X+1)$ indexing the product characteristics and $k = 1, \ldots, K$ indexing the demographic characteristics. As such, the heterogeneous component of the home sharing is

$$\lambda_{i,q} = \sum_b \big(\sum_k \pi_k^b z_{i,k} \big) X_q^b \tag{C.9}$$

where $z_{i,k} \sim P^*_{D_q}$ is drawn from the empirical distribution of the demographics in market q. With logit error, the market share S_q is thus computed as

$$S_q(\delta_q, \theta_2; X_q, D_q) = \frac{1}{N_q} \sum_q P_{i,q}(\delta_q, \theta_2; X_q, D_q) = \frac{1}{N_q} \sum_q \frac{\exp(\delta_q + \lambda_{i,q})}{1 + \exp(\delta_q + \lambda_{i,q})} \tag{C.10}$$

C.2 Analytical Derivations

Analytical Derivatives of MPEC

The gradient of the objective function is

$$\nabla_{(\delta,\theta_1,\theta_2,\eta)} G(\eta; W) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 2W\eta \end{bmatrix} \tag{C.11}$$

The Jacobian for the constraints is

$$\nabla_{(\delta,\theta_1,\theta_2,\eta)} (f, g) = \begin{bmatrix} \frac{\partial f}{\partial \delta} & 0 & \frac{\partial f}{\partial \theta_2} & 0 \\ \\ \frac{\partial g}{\partial \delta} & \frac{\partial g}{\partial \theta_1} & 0 & I_g \end{bmatrix} \tag{C.12}$$

where

$$\frac{\partial f_q}{\partial \delta_q} = \frac{1}{N_q} \sum_i P_{i,q}(1 - P_{i,q}), \quad \frac{\partial f_q}{\partial \delta_q'} = 0, \quad q \neq q' \tag{C.13}$$

$$\frac{\partial f_q}{\partial \pi_k^b} = \frac{1}{N_q} \sum_i P_{i,q}(1 - P_{i,q}) z_{i,k} X_q^b \tag{C.14}$$

$$\frac{\partial g}{\partial \delta} = -Z^T, \quad \frac{\partial g}{\partial \theta_1} = Z^T X \tag{C.15}$$

Note that the upper-left $Q \times Q$ block of $\frac{\partial f}{\partial \delta}$ contains only diagonal terms $\frac{\partial f_q}{\partial \delta_q}$. When the number of markets Q is large, this results in a sparse Jacobian, which is particularly attractive computationally.

Analytical Hessians of MPEC Lagrangian

Next, I derive the analytical Hessian of the MPEC Lagrangian.

$$\nabla^2\mathcal{L}(\delta,\theta_1,\theta_2,\eta) = \begin{bmatrix} \frac{\partial^2\mathcal{L}}{\partial\delta^2} & 0 & \frac{\partial^2\mathcal{L}}{\partial\delta\partial\theta_2} & 0 \\ 0 & 0 & 0 & 0 \\ \frac{\partial^2\mathcal{L}}{\partial\theta_2\partial\delta} & 0 & \frac{\partial^2\mathcal{L}}{\partial\theta_2^2} & 0 \\ 0 & 0 & 0 & \frac{\partial^2\mathcal{L}}{\partial\eta^2} \end{bmatrix} \tag{C.16}$$

where

$$\frac{\partial^2\mathcal{L}}{\partial\delta_q^2} = \lambda_f^q\frac{\partial f_q^2}{\partial\delta_q^2} = \lambda_f^q\frac{1}{N_q}P_{i,q}(1-P_{i,q})(1-2P_{i,q}) \tag{C.17}$$

$$\frac{\partial^2\mathcal{L}}{\partial\delta_q\delta_q'} = 0, \quad q \neq q' \tag{C.18}$$

$$\frac{\partial^2\mathcal{L}}{\partial\delta_q\partial\pi_k^b} = \lambda_f^q\frac{\partial f_q^2}{\partial\delta_q\partial\pi_k^b} = \lambda_f^q\frac{1}{N_q}P_{i,q}(1-P_{i,q})(1-2P_{i,q})z_{i,k}X_q^b \tag{C.19}$$

$$\frac{\partial^2\mathcal{L}}{\partial\pi_k^b\partial\pi_{k'}^{b'}} = \sum_q\lambda_f^q\frac{\partial f_q^2}{\partial\pi_k^b\partial\pi_{k'}^{b'}} = \sum_q\lambda_f^q\frac{1}{N_q}P_{i,q}(1-P_{i,q})(1-2P_{i,q})z_{i,k}X_q^b z_{i,k'}X_q^{b'} \tag{C.20}$$

$$\frac{\partial^2\mathcal{L}}{\partial\eta^2} = 2W \tag{C.21}$$

Again, notice the upper-left $Q \times Q$ block representing $\frac{\partial^2\mathcal{L}}{\partial\delta^2}$ has non-zero entries only on the diagonal term. Thus, even with a large Q, the Hessian remains sparse, making it computationally more tractable.