Estimating Preferences for Neighborhood Amenities Under Imperfect Information *

Fernando Ferreira†
University of Pennsylvania and NBER

Maisy Wong‡
University of Pennsylvania and NBER

May 2022

Abstract

We develop a generalized neighborhood choice model allowing for heterogeneity in knowledge about amenities. When people sort into neighborhoods under imperfect information, all amenities are potentially endogenous. We construct a latent quality index to address this bias, using panel data from a neighborhood choice program that provided information about rents and same-school network to graduating students. Individuals switch into neighborhoods with larger networks and lower rents, and the effects persisted after graduation, influencing actual residential choices. Marginal willingness-to-pay estimates are \$123 per month to live with a larger network, and not accounting for endogeneity will bias it up by 70%.

Keywords: Neighborhood Choice, Residential Sorting, Imperfect Information, Housing Demand, Amenities

*We are grateful for support from the Dean's Research Fund, the Research Sponsors Program of the Zell/Lurie Real Estate Center, and the Wharton Behavioral Lab. We thank Ricardo Perez-Truglia and Nicolas Bottan for sharing their survey design. We thank Eduardo Morales, Gregorio Caetano, Stijn Van Nieuwerburgh, Milena Almagro Garcia, Robert Inman, and participants at seminars in AREUEA, Google, McGill, OSUS, Toronto, Federal Reserve Board/George Washington University, and Wharton, as well as conferences including UEA, the Real Estate NBER Summer Institute, and AEA Meetings for helpful comments and suggestions. Qize Chen, Richard Jin, Rebecca Jorgensen, Jacob Krimmel, Xuequan Peng, Trevor Wooley, Fan Xia, Shuo Yang, and Rui Yu provided excellent research assistance. All errors are our own.

†Corresponding author. Wharton Real Estate. 430 Vance Hall, 3733 Spruce Street, Philadelphia, PA 19104-6301. Email: fferreir@wharton.upenn.edu.

‡Wharton Real Estate. 434 Vance Hall, 3733 Spruce Street, Philadelphia, PA 19104-6301. Email: maisy@wharton.upenn.edu.

1 Introduction

Choosing a neighborhood is an important and complex decision. Several neighborhood amenities are critical for the well-being of adults and children, such as exposure to the social and professional network, and access to public and private goods, among others.[1] But choosing a neighborhood in which to live can be a challenging process given the many neighborhoods available in a labor market area, with each neighborhood having an exceedingly large number of characteristics. In practice individuals have varying knowledge about neighborhoods and their multitude of amenities, and face difficulties processing the terabytes of information available on public websites, such as Zillow. These problems may be even more salient for young adults, since they do not have much experience with residential choices, and may rely on social norms and a limited network of friends to gather information.[2]

In this paper we provide a new framework for estimating the value of neighborhood amenities, accounting for heterogeneity in how much individuals know about a neighborhood. A standard assumption in choice models is that individuals have perfect information about their choice set characteristics – all neighborhoods and their amenities in our case.[3] We present a more general model that allows for heterogeneity in individual knowledge about neighborhood amenities, i.e., some individuals may have imperfect information about the cost of living and school quality in a neighborhood, while others may have imperfect information about demographic composition or the presence of trees and sidewalks in a different neighborhood.

We show that such individual imperfect information leads to a new type of bias in the estimation of preferences for neighborhood amenities. For example, the price elasticity

[1] A few examples from this literature include the value heterogeneous households place on endogenous neighborhood amenities, such as school quality and sociodemographic composition (Bayer, Ferreira, and McMillan, 2007; Wong, 2013), how informal hiring networks arise from place of residence (Bayer, Ross, and Topa, 2008), the importance of private consumption goods (Allcott, Diamond, Dubé, Handbury, Rahkovsky, and Schnell, 2019), negative consequences of local disamenities such as pollution (Chay and Greenstone, 2003, 2005; Isen, Rossin-Slater, and Walker, 2017) and crime (Linden and Rockoff, 2008), and how neighborhoods impact long-term outcomes for children (Chetty, Hendren, and Katz, 2016).

[2] There were almost 20 million individuals enrolled in college in 2019 in the United States (National Center for Education Statistics, 2020), who will soon have to choose where to live. Recent research has highlighted the role of housing costs, amenities, social and professional networks in shaping the location choices of young adults (Moretti, 2013; Diamond, 2016) and the importance of friends in their decision-making (Bailey, Cao, Kuchler, Stroebel, and Wong, 2018).

[3] Discrete choice models applied to neighborhood choice have roots in the work of McFadden (1978); Berry (1994); Berry, Levinsohn, and Pakes (1995, 2004); Epple and Sieg (1999); Nevo (2001); Petrin (2002); Train (2009); Pakes, Porter, Ho, and Ishii (2015). Perfect information is a standard assumption in all these models.

of demand may be upward or downward biased because individuals may under or over estimate cost of living in certain places. Such problems arise not only with prices, but also with all other dimensions of neighborhood quality. In practice, we demonstrate that when individuals sort into neighborhoods under imperfect information about amenities, this can potentially make all amenities endogenous.

In this context, standard methods to account for unobserved neighborhood quality using neighborhood fixed effects or market share inversions (Berry, 1994; Berry, Levinsohn, and Pakes, 1995) can still lead to biased estimates. Our solution makes use of a latent quality index strategy that can account for both average unobserved neighborhood quality and individual imperfect information about amenities. We show in section 2 that the validity of our strategy relies on the assumption that such an index exhausts all the information about how unobserved heterogeneity influences choices.[4] Moreover, we connect this to standard relevance and excludability conditions that motivate a set of empirical assessments we use to characterize the key sources of identifying variation underpinning our structural estimation. This latent index strategy is also flexible and can be easily generalized to other settings.[5]

To estimate our generalized neighborhood choice model we use data from a choice program that allows us to observe individual rankings before and after an information intervention. We partnered with a large professional school in the East Coast to develop a neighborhood choice program to help graduating students choose where to live. This is a unique and consequential setting due to the large number of movers and the fact that most students in the sample already knew their place of work and were highly motivated to learn more about neighborhood options. The school was worried about students having imperfect information about neighborhood quality, and students had concerns about cost of living and access to professional and social networks. Given those issues, the program focused on providing information to all students about two neighborhood characteristics: average rent and the same-school network shares.

The neighborhood choice program was launched by the school in April 2019 for students who were in the early stages of their housing search. The first part of the program was a six-minute survey with four main components: 1) Students ranked neighborhoods in

[4] For a general discussion see Berry and Haile (2014) and Dahl (2002).

[5] Examples include research using panel data on location decisions or individuals facing different information sets. Potential applications include choice of college major (Wiswall and Zafar, 2014), choice of cities by medical residents Bottan and Perez-Truglia (2022), choice induced by rent control (Diamond et al., 2019), and by entry of big firms (Qian, 2021).

their chosen Metropolitan Area (MSA); 2) Students indicated their best estimates of rents and network shares in those neighborhoods; 3) We provided information about market rents (from Zillow) and network shares (from administrative data) for all neighborhoods; 4) Students were asked to rank neighborhoods again. The second part of the program, received upon completion of the survey, was an interactive map with granular information about rents and the same-school network. Students had permanent access to the map, and could use that information during their summer housing search and final neighborhood choice.

A total of 341 students completed the survey (40% response rate), with 309 students choosing one of our 20 available MSAs.[6] For each MSA students had a menu with an average of 19 neighborhoods, and they could rank up to 10 of their preferred neighborhoods – an average of five neighborhoods were actually ranked. Survey responses showed significant heterogeneity in their initial knowledge of neighborhood amenities. On average, students under-estimated monthly rent by \$620, and over-estimated network shares by six percentage points. There is also asymmetric heterogeneity with students over-estimating rents by \$140 for neighborhoods below \$2,500 and under-estimating rents by more than \$1,000 for neighborhoods above \$4,000. Students also tended to over-estimate network shares by more in expensive neighborhoods, consistent with the administrators' concerns that students had imperfect information.

Comparing rankings pre and post information reveal that individuals systematically prefer expensive neighborhoods with better amenities. In particular, neighborhoods that are always ranked in the top three for a given student tend to have higher rents and network shares relative to neighborhoods that are never ranked in the top three.

Crucially, we also observe switchers who change their rankings after the information intervention in a way that increases the network shares by 1.46 percentage points and decreases rents by \$430 for their top neighborhoods. We show that such switchers are representative of the overall sample, and that the findings are robust to controlling for individual fixed effects. These switching patterns suggest a negative marginal utility for rent, and a positive marginal willingness to pay to live close to a larger network. The variation from switchers is critical since our estimation strategy relies on changes in rankings and students who exhibit persistent tastes will tend to have collinear pre and post information rankings.

Estimation of our neighborhood choice model under imperfect information follows two stages. In the first stage we recover heterogeneity in preferences by estimating a rank ordered logit model where post information neighborhood rankings are a function of market

[6] In the past decade, 67% of students chose to move to these 20 MSA's upon graduation.

rents, actual same-school network shares, and neighborhood-by-individual unobservables. The first stage also produces estimated neighborhood fixed effects that are decomposed into mean preferences in a second stage. Individuals have their own unique consideration sets, based on their selected neighborhoods, and we allow for ample heterogeneity in preferences by age, gender, first-generation and minority groups, marital status, school major, industry, and location of prior residence.

To account for amenity endogeneity, we use pre information rankings. The idea is that individual level pre-information rankings can be used as latent quality indices that summarize all individual knowledge about neighborhood quality. The validity of the indices relies on two assumptions. First, individuals report pre information rankings that remain relevant proxies for the desirability of neighborhoods. This is plausible as long as preferences remain stable in the two minutes between pre and post information rankings (Wiswall and Zafar (2014)). And second, changes in rankings after the information intervention reflect new information about the networks and rents, i.e., there are switchers who respond to the information intervention. A potential threat to identification occurs if students change their views about unobserved amenities once they receive new information about neighborhood prices. We assess this threat by comparing student's views of a third variable, the walk score of a neighborhood. We find that students barely change their walk score views after receiving information about rents and networks, and that such small change is uncorrelated with rents and networks. Another potential issue is that switchers might be a selected sample, but we show in the empirical work that they are representative of the student body.

Our generalized model with latent quality indices estimates a marginal willingness to pay of \$123 per month for a 1 percentage point increase in network shares, given an average monthly rent of \$3,600 in our estimation sample. This is 70% lower than a model without latent quality indices (MWTP of \$400). This reduction in marginal willingness to pay is mainly driven by a more negative and precise estimate of the marginal utility for rents, while preferences for the network remain somewhat constant. This is consistent with the nature of imperfect information that gives rise to a positive correlation between rent and latent preferences for neighborhood quality. That in turn biases the marginal utility of rent towards zero, leading to inconsistent estimates of MWTP for amenities. Such biases are addressed by our latent quality index.

Next, we demonstrate the flexibility of our empirical strategy by constructing different latent quality indices to empirically assess the varying manners in which unobserved heterogeneity can bias our MWTP estimates. These robustness tests confirm the internal validity of our main estimates. We find that just adding neighborhood-level averages of pre information rankings can reduce the MWTP by 61% to \$157, while just including the individual-level rankings reduces the MWTP by 54% to \$184. Including both is needed to reduce the MWTP to \$123. Thereafter, the MWTP remains stable even if we augmented the model with additional proxies using initial knowledge about rents and network shares and post-information knowledge about walk scores. We also use our survey data to better understand why individuals prefer to live in a neighborhood with a larger same-school network. Social and professional interactions account for the largest fraction of the responses. Moreover, we include richer measures of individual heterogeneity using administrative data, including industry, major, stage of search, but none of them lead to significant changes in the main MWTP estimate.

Finally, we demonstrate that the patterns in the online survey persist up to a year after the program by collecting data on how individuals searched for neighborhoods and their neighborhood choices after graduation. Our measure of search is based on neighborhood clicks from the interactive maps made available to students upon the survey completion. We also obtained administrative data on actual neighborhood choices after graduation. We find the same behavior for switchers, i.e., students tend to switch into neighborhoods with lower rents and larger network, while switching out of neighborhoods with higher rent and smaller network. These results indicate that the new neighborhood choice program successfully provided consistent and systematic information that allowed students to improve their location decisions. Indeed, they switched into neighborhoods where average monthly rents are \$532 lower, implying significant cost savings relative to the average monthly rent of \$3,371 based on where they live. Network shares remained similar. These results also confirm that marginal utility estimates from the neighborhood choice model based on stated preferences, translated into changes in revealed neighborhood preferences.

To summarize, we make five contributions to the literature on neighborhood sorting models. First, we generalize the standard neighborhood choice model to allow for unobserved heterogeneity in imperfect information about neighborhood amenities. This is likely pervasive in the context of neighborhood choices, and it gives rise to a new source of omitted variable bias affecting all neighborhood characteristics. Second, we overcome this issue by developing a latent quality index that deals with biases associated with both average unobserved neighborhood quality and imperfect information. We establish intuitive relevance and excludability conditions that can be empirically assessed transparently. Third, we estimate the model using detailed micro data from a set of movers that are about to face a consequential decision of choosing a neighborhood to live upon graduation. Fourth, our data provides rankings of neighborhoods under different information sets, and the presence of switchers who change their rankings when information changes can help with the empirical identification. Fifth, we reinforce our survey findings using administrative data that allows us to follow individuals post graduation. Reassuringly, the targeted information interventions have persistent effects up to one year after the survey, translating into significant savings in cost of living.

Other recent work has made progress in modeling neighborhood choice using observational data and structural assumptions in different ways, such as Bayer, Keohane, and Timmins (2009) who account for moving costs, Bayer, McMillan, Murphy, and Timmins (2016) who consider neighborhood choice in a dynamic framework, Calder-Wang (2019) who allows for changes in the supply of houses in a neighborhood, Caetano and Maheshri (2019) who allow for choices to be observed out of equilibrium, Büchel, Ehrlich, Puga, and Viladecans-Marsal (2020) who utilize high frequency mobile phone data to understand spatial mobility patterns, and Almagro and Domínguez-Iino (2020) who allows for changes in the supply of amenities in neighborhoods. Our work also complements recent methodological advances in discrete choice models that address selection bias using hedonic indices (Epple, Quintero, and Sieg, 2020) and fixed effects (Pakes and Porter, 2016; Honoré and Hu, 2020).

Our research is also part of a growing body of work that combines surveys, experiments, and structural estimation, such as Benjamin, Heffetz, Kimball, and Rees-Jones (2014) who study life satisfaction of students who submit choice rankings to medical schools, and Galiani, Murphy, and Pantano (2015) who use the moving to opportunity experiment to simulate the effect of changes in the subsidies, Bottan and Perez-Truglia (2022) who apply information surveys to medical students in order to understand the importance of relative income in city-level choices, and Bergman, Chan, and Kapor (2020) who study the role of imperfect information about school quality.

There is a theoretical literature on different dimensions of imperfect information, such as Kacperczyk, Van Nieuwerburgh, and Veldkamp (2016) and Gao, Sockin, and Xiong (2021), and applications to household mobility, such as Fujiwara, Morales, and Porcher (2019) and Kosar, Ransom, and Van der Klaauw (2021). Finally, our work is related to programs that provide information and assistance for low income families to move to better neighborhoods, such as Kling, Liebman, and Katz (2007) and Bergman, Chetty, DeLuca, Hendren, Katz, and Palmer (2020).

The paper proceeds as follows: In section 2 we describe our general neighborhood choice model, and we present the survey design in section 3 and descriptive analysis in section 4. In section 5 we explain our estimation approach, and results are shown in section 6. Section 7 concludes the paper.

2 Neighborhood Choice Model

We model the neighborhood location decision of each individual as a discrete choice, following the utility function specification of the random utility models originally developed by McFadden (1973, 1978) and Berry, Levinsohn, and Pakes (1995). The individual i's indirect utility from choosing neighborhood j among J_m neighborhoods in labor market area m is:

$$u_{ijm} = \omega_{jm}\beta_i^c + \varepsilon_{ijm} \tag{1}$$

where ω_{jm} is a C-dimensional vector that includes all neighborhood amenities, β_i^c is a vector of individual preferences for each neighborhood amenity, and ε_{ijm} is an i.i.d. Type-I extreme value error term that reflects i's idiosyncratic preference for neighborhood j. For simplicity, we will suppress the subscript m for now. Preferences for each amenity c are a function of the individual's own observed demographic attributes z_{id}:

$$\beta_i^c = \beta_o^c + \sum_{d=1}^{D} \beta_d^c z_{id} = \beta_o^c + \beta_{id}^c \tag{2}$$

The vector ω_j literally includes all characteristics that define a neighborhood, such as price, quality of the housing stock, demographic composition, number and type of trees, air quality, crime rates, number and type of restaurants and bars, access to sidewalks, etc.

However, given the high dimensionality of neighborhood amenities, imperfect information is likely pervasive and individuals may observe a large set of amenities with error. We augment equation 1 as follows:

$$u_{ij} = \omega_j\beta_i^c + \Delta\omega_{ij}\beta_i^c + \varepsilon_{ij} \tag{3}$$

where $\Delta\omega_{ij}$ captures a flexible notion of imperfect information, allowing for some individuals to have imperfect information about the cost of living in a neighborhood, while others may have imperfect information about the presence of trees and sidewalks in other neighborhoods.[7] Our generalized model nests perfect information as a special case. If individuals have common and complete knowledge, $\Delta\omega_{ij}$ will be null vectors for all individuals and all neighborhoods.

In practice econometricians do not observe all characteristics ω^c, and in general it is not feasible to estimate choice models with thousands of characteristics. Assuming the econometrician observes a limited number of characteristics x^k:

$$u_{ij} = x_j^k \beta_i^k + x_j^{k^-} \beta_i^{k^-} + \Delta\omega_{ij}\beta_i^c + \varepsilon_{ij} \tag{4}$$

The new term $x_j^{k^-}$ for neighborhood j represents the $c-k$ neighborhood characteristics that are unobserved by the econometrician. In models with perfect information the key source of bias in the estimation of β_i^k is that observed neighborhood characteristics (x^k) are generally correlated with the unobserved neighborhood quality in $x_j^{k^-}$. To address this identification problem, one can control for unobserved neighborhood quality using a combination of neighborhood fixed effects - which can be estimated using panel data or market share inversions (Berry, 1994) - and neighborhood shifters.[8] However, once we allow for imperfect information, the identification problem cannot be solved using standard methods alone, since they generally account only for the omitted factor $x_j^{k^-}$ but not necessarily for endogeneity associated with $\Delta\omega_{ij}$. Moreover, all variables in x_j^k become endogenous when allowing for this generalized form of imperfect information.

To simplify our notation, we collect the terms in equation 4 that are unobserved by the econometrician as ξ_{ij}. This captures both unobserved neighborhood quality $x_j^{k^-}$ and also heterogeneity in knowledge of amenities, $\Delta\omega_{ij}$.

$$u_{ij} = x_j^k \beta_i^k + \xi_{ij} + \varepsilon_{ij} = V_{ij} + \varepsilon_{ij} \tag{5}$$

[7] This setting can also allow for individuals to have different expectations over future amenities. Expectations are less relevant in our empirical context though, since the majority of students will rent apartments for one or two years as opposed to buying homes. See Bayer, McMillan, Murphy, and Timmins (2016) for a more structural approach to model evolving future local amenities.

[8] This strategy involves estimating the relationship between estimated fixed effects and average neighborhood characteristics in a second stage, where instrumental variables are needed in order to deal with the correlation between observed characteristics and unobserved quality. See for example Bayer, Ferreira, and McMillan (2007).

2.1 Identification under Imperfect Information

To identify β_i^k, we propose to construct proxies of ξ_{ij} using a latent quality index, $g(\widetilde{\xi}_{ij})$. Let F be the joint distribution for utilities u_{ij}, conditional on x_j^k and ξ_{ij}. Then, $g(\widetilde{\xi}_{ij})$ satisfies the index sufficiency assumption if:

$$\textit{Index sufficiency:} \qquad F_u[u_{i1},\cdots,u_{iJ}|x_j^k,\xi_{ij}] = F_u[u_{i1},\cdots,u_{iJ}|x_j^k,g(\widetilde{\xi}_{ij})] \tag{6}$$

The index sufficiency assumption is satisfied if $g(\widetilde{\xi}_{ij})$ exhausts all the information about how ξ_{ij} influences neighborhood choices in equation 5. This characterization of index sufficiency is similar to Dahl (2002) and Manski (1988), and we follow Berry and Haile (2014) in allowing the single index to enter the utility in the same way as other observed features. In section V we will show that, empirically, such index will be a function of pre-information rankings from the same individual.

To develop intuition, consider a researcher estimating the MWTP for a neighborhood amenity, A. For simplicity, assume utility follows a version of equation 5 where neighborhood utility can be characterized using observed integer ranks, R_{ij}, and all neighborhoods are differentiated along three dimensions only: cost of living P_j, amenity A_j, and the number of trees T_j (unobserved). For the baseline case, individual information is imperfect for all three amenities, so that utility depends on the actual amenity levels and also heterogeneity in knowledge:

$$R_{ij}^0 = (P_j + \Delta P_{ij}^0)\beta_i^P + (A_j + \Delta A_{ij}^0)\beta_i^A + (T_j + \Delta T_{ij}^0)\beta_i^T + \varepsilon_{ij}^0 \tag{7}$$

where the terms in gray are unobserved by the econometrician (ξ_{ij} in equation 5). There are a host of omitted variables that confound the estimate of β_i^P, including the standard quality confounder whereby expensive neighborhoods have more trees ($Cov(P,T) > 0$), under-estimation of the cost of living in expensive neighborhoods by some individuals ($Cov(P,\Delta P) < 0$), or over-estimation by other individuals ($Cov(P,\Delta P) > 0$). Additionally, some individuals may use price as a heuristic of quality, thereby over-estimating A and T in expensive neighborhoods ($Cov(P,\Delta A) > 0$ and $Cov(P,\Delta T) > 0$). The identification challenges for β_i^A are analogous. This simple example illustrates how identification can be challenging since it depends on the nature of the imperfect information for each individual. The multi-dimensional nature of $\Delta\omega_{ij}$ gives rise to many potential correlates with P and A.

Next, consider individuals facing a different information set where everyone has common and complete knowledge about P and A so that $\Delta P_{ij} = 0$ and $\Delta A_{ij} = 0$.[9]. In our empirical application below we will use data from a program that provided information about prices and amenities to individuals. Under this new information environment, R_{ij} will now be:

$$R^1_{ij} = P_j\beta^P_i + A_j\beta^A_i + (T_j + \Delta T^1_{ij})\beta^T_i + \varepsilon^1_{ij} \tag{8}$$

where ΔT^1_{ij} represents unobserved heterogeneity in knowledge of T_j and ε^1_{ij} reflects i's idiosyncratic taste for j under this new information set. Notably, even though there is complete knowledge of P and A, there could still be confounding due to $T_j + \Delta T^1_{ij}$.

We propose to use R^0_{ij} to construct a latent quality index that can serve as a proxy of quality in the new information environment. Formally, $g(R^0_{ij})$ satisfies the index sufficiency assumption if $F_R[R^1_{i1}, \cdots, R^1_{iJ}|P_j, A_j, T_j + \Delta T^1_{ij}] = F_R[R^1_{i1}, \cdots, R^1_{iJ}|P_j, A_j, g(R^0_{ij})]$.

Relevance and Exclusion Restriction. The credibility of the latent quality index can be understood by looking at relevance and exclusion restrictions associated with index sufficiency. For $g(R^0_{ij})$ to exhaust all the possible ways in which $T_j + \Delta T^1_{ij}$ influences R^1_{ij}, it would be useful if changes in rankings are only due to shifts in the information environment around P and A ($\Delta P^0_{ij} \neq \Delta P^1_{ij}$ and $\Delta A^0_{ij} \neq \Delta A^1_{ij}$) whilst knowledge for other amenities remain comparable across the two information settings ($\Delta T^0_{ij} = \Delta T^1_{ij}$). The latter assumption also suggests a natural relevance condition whereby R^0_{ij} can serve as a proxy of unobserved neighborhood quality for R^1_{ij}, above and beyond conditioning on P and A.

Comparing equations (7) and (8) also highlights the need for exogenous shifters of information for P and A as identifying sources of variation. Since neighborhood preferences tend to be stable, R^1_{ij} and R^0_{ij} are likely highly collinear. Below, we propose to use data from a new neighborhood choice program that allows us to observe individual rankings of neighborhood choices under two different information sets.

[9]We assume complete knowledge in this stylized example for exposition. We relax this assumption in our estimation and provide tests in Table 9.

3 Information Survey and Data

3.1 Neighborhood Choice Program

We partnered with a large professional school in the East Coast to design a neighborhood choice program to help students choose where to live after graduation. In our discussions with students and administrators to understand how students chose neighborhoods, many acknowledged that this was a complex decision given the large number of neighborhood characteristics and the large number of neighborhoods to choose from. Four main issues surfaced in qualitative interviews. First, students mentioned anxiety about cost of living due to high housing costs in many cities. Second, students highly value access to the professional and social network of fellow students and alums, and wanted to preserve that network after graduation (Shue, 2013). Third, students had unequal access to information and relied on limited networks to obtain neighborhood information.[10] Finally, students believed in a social norm where the same-school network tended to live in neighborhoods with high cost of living, inadvertently leading students to choose expensive neighborhoods upon graduation. To address these concerns, we designed a program to provide all students with information about cost of living and the same-school network shares, to help them choose neighborhoods.

In April 2019, the Vice Dean of the school emailed all students from the graduating cohort to introduce the new neighborhood choice program. April is the ideal timing because it is about a month before graduation and a majority of the students already have a job and know which city they want to move to - and at the same time most students had just started the process of searching for housing in their new destination. The program turned out to be an important tool for students who were about to make a very consequential choice for their lives. The program provided neighborhood information in two ways. Students would first access an online survey which provided information about the neighborhoods in their preferred metropolitan area in the United States and also asked basic questions about their neighborhood choices. Students were given a $25 Amazon gift card to encourage them to complete the survey. After completing the survey, students could access a mapping tool which provides the same information at a more granular spatial resolution for all metropolitan areas available in our data.

[10] In our survey, 88% of students report speaking to fewer than four contacts about their search process, 95% connected with fewer than four contacts online or through social media.

3.2 Neighborhood Information

Below, we describe how we defined neighborhoods, and explain the type of information provided to survey takers.

Neighborhood names. We begin by selecting the top 20 most popular labor markets (MSAs) in the United States, based on the current residence of all school cohorts who graduated from 2010 to 2018. Other MSAs with a small number of graduates were not used in order to preserve data confidentiality. We then split each MSA into a set of comprehensive yet parsimonious neighborhoods. As a baseline, we used shapefiles from Zillow which classifies the urban core into neighborhoods. In places without Zillow neighborhoods - usually suburban areas - each county would be a neighborhood. In some instances where we had to combine neighborhoods to reduce the total number of choices in each MSA, we joined adjacent neighborhoods with similar levels of college graduates, based on the census, and reported both names in the survey. To generate a list of neighborhood names that students would be familiar with, we relied on Google Trends data. In particular, when there were multiple ways to identify a location, we chose the most popular name according to Google Trends. Ultimately, we ended up with an average of 18.5 neighborhoods across all MSA's. Columns 1 and 2 of Table 1 lists the MSA's and the number of neighborhoods in each MSA.

Cost of Living. Cost of living was a major concern since many students have student loans and live in cities with high housing costs. We focus on rental housing costs as the main metric for cost of living. Monthly rents were preferred over housing prices because the vast majority of students occupy rented residences in their first few years after graduation. We obtained monthly rents from Zillow, which publicly provides a monthly rent index for an average home in each neighborhood. We chose to present information about the average 2018 rent using all months in order to mitigate outliers coming from solely using one or two months of data.

Same-School Network. To describe the same-school network comprehensively, we first obtained proprietary administrative data with the current street addresses of all recent graduates of the school. The school utilizes various sources to ensure that the addresses are current and accurate. We focused on the cohorts who graduated between 2010 and 2018 - this aggregation was required in order to preserve student privacy. Additionally, survey respondents do not have access to the total number of individuals living in individual neighborhoods. Instead, we present them with same-school network shares in neighborhood j in

Table 1: Number of Neighborhoods and Percent of Respondents by MSA

MSA	Number of Neighborhoods	Percent of Respondents
Atlanta	19	2.3
Austin	14	0.6
Baltimore	19	0.3
Boston	21	5.5
Chicago	22	4.2
Washington, DC	20	4.9
Dallas	18	1.6
Denver	19	1.0
Houston	15	0.6
Los Angeles	22	2.9
Miami	20	0.6
Minneapolis	17	0.3
New York	25	43.7
Philadelphia	20	6.1
San Diego	20	0.3
San Francisco	22	20.4
San Jose	18	1.0
Seattle	22	3.6
Total	353	100.0

Notes: Top 18 MSA's in the neighborhood choice program with the number of neighborhoods and percent of respondents for each MSA. We offered 20 MSA's but Bridgeport (10 neighborhoods) and Trenton (7 neighborhoods) were not chosen by any student.

MSA m (N_{jm}), by dividing the total number of graduates currently living in neighborhood j by the total number of graduates living in the MSA.

3.3 Survey Design

We designed the neighborhood choice survey to collect unique information about how individuals make choices before and after receiving information about neighborhoods. Each student would first choose a metropolitan area in which they were planning to live upon graduation. They had to choose among 20 MSAs or select the option "None of the above". In this latter case students self-reported the name of another city in the US or abroad - 32 students picked that option and we exclude them from the remaining analysis. Another 309 students chose one of the top 20 MSAs and completed the survey. Column 3 of Table

1 shows that forty four percent of these respondents selected New York, followed by San Francisco (20%), Philadelphia (6%), Boston (6%) and DC (5%).

Around 40% of the graduating cohort of students participated in the neighborhood choice program. Panel A of Table A1 in the Appendix reports the summary statistics of our respondents. The average age is 29, half of them are female, 15% are first-generation or part of an under-represented minority, 12% are married or have children, and 23% are international students (by citizenship). We also show that average characteristics of survey takers are not that different from non-respondents, except respondents are less likely to be international students since they are less likely to want to live in the United States.[11]

Then we asked students to rank up to ten neighborhoods in that metropolitan area, allowing them to create their own neighborhood consideration sets. Only for these neighborhoods we then ask students to provide their best estimates of the same-school network shares and the monthly rent for the average home. Next we displayed the program information about the rent and network shares in *all* neighborhoods in that MSA.[12] Here, each respondent would also see in a figure how this new information compared with her own unique estimates.

After presenting the information, we asked students to re-rank up to ten neighborhoods. This page looks identical to the pre information stage, except with the new information just so students would not need to scroll back in order to check the data. Students could choose from a menu of all the neighborhoods along with information about the rent and network shares. We did not pre populate this page with their prior rankings so as not to prime their post information choices. Finally, at the end of the survey we also asked some questions about whether and why they thought the information influenced their neighborhood choices, and other factors related to their neighborhood search processes. The Survey Appendix provides more details about the survey questions.

The survey was designed to be short. The median student completed the survey in six minutes, spending one minute on the pre information ranking of neighborhoods, seventy seconds to estimate the rent and network shares in their chosen neighborhoods, forty

[11]Indeed, the coefficient on the international student dummy falls from 0.17 to 0.09 when comparing the 309 respondents who chose the top-20 MSA's in our neighborhood choice program (Panel A of Table A1) versus the full set of respondents, including the 32 who chose other cities (Panel B). These demographics are based on administrative data from the school. We also observe other pre-determined characteristics, such as, their industry, intended major, where they worked before attending school.

[12]We randomly assigned the order students would receive the information about rents and network shares, in order to potentially test salience effects. This order did not have any impact on our results described below, indicating that salience effects may not be an issue.

Table 2: Summary Statistics for All and Considered Neighborhoods

	All			**Considered (pre)**			**Considered (post)**		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Same-School Network Share	353	5.10	5.01	213	6.28	3.89	193	6.89	4.04
Rent (thousands)	353	2.47	1.01	213	3.65	1.28	193	3.60	1.22
Income (thousands)	353	80.26	25.45	213	98.28	31.21	193	97.59	29.10
Bachelor′s Degree+	353	0.46	0.19	213	0.65	0.17	193	0.65	0.16
Minority Share	353	0.38	0.19	213	0.38	0.15	193	0.37	0.15

Notes: Summary statistics for amenities in all (353) neighborhoods in the neighborhood choice program, as well as the 213 (193) neighborhoods considered pre (post) information. The five amenities include the same-school network share, average monthly rent from Zillow, average income, share of population with a college degree or more, non-White share. The latter three are from the 2010 Census.

seconds to read about the rent and network shares for the full set of neighborhoods, and another one minute to re-rank neighborhoods.

The neighborhood choice program did not include a control group since the objective of the program was to promote equal access to information. Having a control group which did not receive the same information intervention would raise equity concerns. Given the nature of the information intervention, it would also be difficult to prevent treated students from sharing information with those assigned to the control group. Moreover, the key identifying variation in our model comes from comparing neighborhood rankings by the same individual before and after the information intervention. Nonetheless, as an additional check, we demonstrate in Table A4 that our findings are robust when we compare location decisions for participants in the program to students in the same cohort who did not participate.

Table 2 presents the neighborhood characteristics of all the 353 neighborhoods, as well as the 213 and 193 neighborhoods considered before and after information, respectively. Interestingly, individuals only rank an average of 5 neighborhoods, even though they can rank up to 10 neighborhoods in each MSA. The chosen neighborhoods have higher monthly rent ($3,600 post information) relative to $2,470 for all neighborhoods. The ranked neighborhoods also have higher network shares (6.89 percent post information) compared to all the neighborhoods (5.1 percent). Moreover, the average income and the college share in the chosen neighborhoods are also higher.

Mapping and Actual Choices. Upon completion of the survey, students were directed to a restricted-access mapping service with the information about rents and network shares at an even more granular geographic resolution for all metropolitan areas available in our data. These interactive maps require students to click in neighborhood geographies in order to access the relevant information. The maps became permanently available in order to help students during their housing and neighborhood search, and we collected data on map clicks. Almost 55% of the students had at least one recorded interaction with the maps from the time of survey completion until graduation. We count the number of clicks students make in each neighborhood as a proxy of how intense they are searching in a neighborhood.

Even though the survey does not formally request information about students' actual neighborhood choices, we obtained post graduation residential locations from the school alumni office. That office utilizes various sources to keep updated location information of alums, and was able to provide accurate addresses for 176 graduates from our sample.[13] We georeferenced them in order to compare where students chose to live after graduation relative to stated preferences in the online survey. Table A2 in the Appendix shows that both samples of students where we have map clicks data and actual post graduation locations are largely representative.

Walk Score. Our exclusion restriction assumes that changes in rankings reflect only new knowledge about rents and networks. This would be violated, for example, if individuals also updated their knowledge about other amenities. The 2019 survey did not include questions about other amenities, but in April 2022 we conducted a new survey for graduating students to learn how neighborhood choices were impacted by the COVID pandemic and the rising prevalence of work-from-home (Ferreira and Wong, 2022).

To shed light on our exclusion restriction, we added a question on individuals' knowledge about the walk score of neighborhoods. We chose the walk score because it is a widely known metric measuring how easy it is to walk to amenities in a neighborhood. It is also an amenity that students value and is likely to be correlated with network shares and rent. We classified the walk score into four discrete levels: Very High (90 to 100), High (70 to 89), Low (50-69), Very Low (0-49). As a benchmark, we provided two well-known neighborhoods near campus with walk scores in the Very High and High range. In our pilots and focus group discussions, we confirmed that students valued walkability and understood the notion of a walk score.

[13]The process to update the location information of alums has been disrupted due to the Covid pandemic. Most of the address information reflect post graduation choices made before the Covid pandemic. As a conservative robustness exercise, our conclusions remain the same using a subset (110 out of 176) of location information identified in January 2020, before the pandemic.

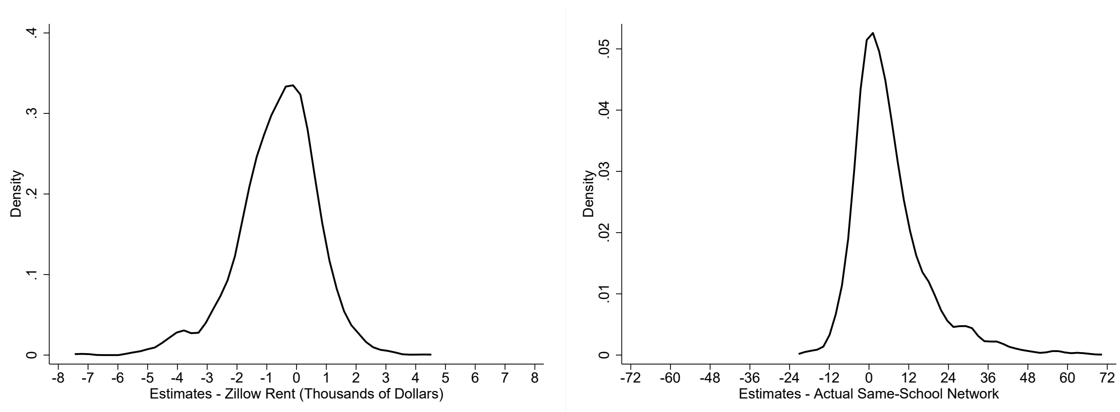
We asked about the walk score twice in the survey. In the pre information stage, we did so after we asked for the best estimates of rent and network shares. In the final question of the survey, we asked students again for the walk score of neighborhoods considered in the post information stage. In total, we have 309 respondents and 1,295 walk scores reported by individual i for neighborhood j before and after information.

4 Descriptive Analysis

4.1 Heterogeneity in Neighborhood Knowledge

Figure 1: Differences Between Reported and Actual Neighborhood Attributes

Panel A: Rent **Panel B:** Same-School Network Shares



Notes: Panel A presents kernel density estimates of the difference of individuals' best guesses of neighborhood rent and Zillow rent for each neighborhood ranked in the pre information period. Panel B repeats the same for same-school network shares.

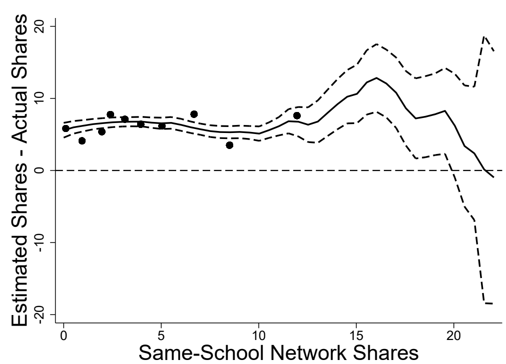
Panel A of Figure 1 presents a kernel density of students' estimates of rents minus the neighborhood rents from Zillow. This figure includes 1,910 neighborhood-by-individual choices where we have rent estimates for neighborhoods considered in the pre period by 309 students. On average, students under-estimate monthly rent by \$620 relative to the average monthly rent of \$3,650. But there is a fair amount of heterogeneity: about two-thirds of the choices are underestimates and about one-third of the choices are over-estimates. Even though students eventually see the actual price of their home, imperfect information around the average cost of living in a neighborhood can be consequential since our students are in early stages of the search process and only consider 5 neighborhoods on average.

Figure 2: Binscatter for Differences in Reported and Actual Amenity

Panel A: Rent Differences vs. Zillow Rent

Panel B: Network Share Differences vs. Zillow Rent

Panel C: Network Share Differences vs. Actual Network Shares



Notes: The solid line in Panel A represents the estimate from a local polynomial regression of the differences in the vertical axis (e.g. individuals' estimate minus Zillow rent) on Zillow rent. The dashed lines correspond to 95% confidence intervals. We trimmed the figure by dropping estimates at the boundaries. The dots correspond to the average within each decile for Zillow rent. Panel B presents the same but using differences in network shares versus Zillow rent. Panel C presents differences in network shares versus network shares.

Interestingly, the differences are correlated with rent levels in that students underestimate rents in expensive neighborhoods and over-estimate rents in cheap ones. Panel A of Figure 2 shows that students over-estimated rents by \$140 for neighborhoods below \$2,500, and under-estimated rents by more than \$1,000 for neighborhoods above \$4,000.

Turning to knowledge about same-school network shares, Panel B of Figure 1 compares estimates and actual network shares for neighborhoods ranked in the pre period. Of the 1,910 estimates of network shares, sixty-nine percent are over-estimates. On average, students over-estimate the network shares by 6 percentage points - a fair amount relative to a mean of 6 percent for all neighborhoods considered in the pre information stage.

Students also appear to over-estimate network shares by more in expensive neighborhoods. In particular, Panel B of Figure 2 shows a positive correlation between the degree of over-estimation and rent levels for neighborhoods with rents above \$4,000. This echoes the concern of school administrators that a social norm whereby same-school alums tended to live in expensive neighborhoods could influence students to only consider a small set of expensive neighborhoods. Panel C in the same figure shows no relationship between the difference in network share estimates and the actual shares.

Overall, the figures above show remarkable heterogeneity around the nature of imperfect information. The differences between students' estimates and the data provided also correlate with rent levels. Such imperfect information could lead to omitted variable biases in the estimation of neighborhood preferences, especially the taste coefficient for rent.

4.2 Rankings Before and After Information

Table 3 presents a simple cross-tabulation to compare network shares and rents for neighborhoods ranked in the top 3 before and after information. The sample includes 7,012 potential individual-by-neighborhood choices. The diagonal entries in Panel A show that individuals systematically prefer high rent neighborhoods, consistent with the common omitted variable problem that neighborhoods that have high rent tend to have high unobserved quality relative to neighborhoods that are never chosen. Specifically, the top left cell reports the average Zillow rent (\$3,627) for neighborhoods that are always ranked in the top 3 before and after information (9% of 7,012 choices were always ranked in the top 3). The bottom right cell reports the average Zillow rent (\$3,007) for neighborhoods that are never ranked in the top 3 (83%).

The key sources of variation we use to identify preferences for amenities rely on

Table 3: Comparing Top Three Neighborhoods Pre and Post Information

Panel A: Rent

		Pre	
		Yes	No
Post	Yes	Always Top 3 $3,627	Switch In $3,380
	No	Switch Out $3,844	Never Top 3 $3,007

Panel B: Same-School Network

		Pre	
		Yes	No
Post	Yes	Always Top 3 7.73%	Switch In 7.18%
	No	Switch Out 5.72%	Never Top 3 3.87%

Notes: Panel A reports the average rent of neighborhoods. The diagonal cells include neighborhoods that are always and never ranked top three by a given individual before and after the information intervention. The off-diagonal cells report average rent for neighborhoods that are switched in or out of the top three for a given individual, after information relative to before. Panel B repeats the same for same-school network shares.

changes in rankings after the information intervention. The off-diagonal cells show that students tended to switch in neighborhoods that have lower average rent ($3,380) and switch out neighborhoods that have higher average rent ($3,844). This is consistent with individuals responding to new information and updating their neighborhood rankings to choose neighborhoods with lower rent. Both types of top 3 switches correspond each to 4% of all individual-by-neighborhood choices.[14]

The pattern for rents also provides empirical support for our exclusion restriction (Section 2.1). Someone who learned that neighborhood j is more expensive would rank it *less* favorably, consistent with the evidence of individuals switching out of expensive neighborhoods. The exclusion restriction fails if individuals also update that j has better amenities, to the extent that price is a signal of quality. However, this will lead individuals to rank j *more* favorably. Since price and quality affect utility in opposite manners, we can conclude

[14] Table A3 in the Appendix show minimal compositional differences along demographics when comparing switchers to non-switchers.

Table 4: Rent and Same-School Network for Switched Neighborhoods

Dependent variable:	**Rent**		**Same-School Network**	
	(1)	(2)	(3)	(4)
Always in top 3	0.68***	0.68***	3.79***	3.78***
	(0.03)	(0.03)	(0.17)	(0.17)
Switch in	0.39***	0.41***	3.30***	3.36***
	(0.05)	(0.05)	(0.25)	(0.26)
Switch out	0.82***	0.84***	1.86***	1.91***
	(0.06)	(0.06)	(0.21)	(0.22)
N	7012	7012	7012	7012
R-squared	0.40	0.41	0.13	0.13
Switch in - Switch out	**-$434**	**-$432**	**1.44**	**1.46**
p-value	**0.00**	**0.00**	**0.00**	**0.00**
MSA FE	Y	N	Y	N
Demographics	Y	N	Y	N
Individual FE	N	Y	N	Y

* 0.10 ** 0.05 *** 0.01

Notes: OLS regressions including the full set of 7,012 neighborhood-by-individual level choices. The dependent variables are monthly Zillow rent in thousands of dollars (columns 1 and 2) and same-school network shares (columns 3 and 4). The key regressors include a dummy that is one for neighborhoods that are always ranked in the top three by a given individual before and after information, as well as neighborhoods that were switched in and switched out of the top three choice set after information, relative to neighborhoods that were never top three (the omitted group). Columns 1 and 3 have MSA fixed effects and demographic controls, including age, and a dummy for female, married or with children, under-represented minority and first-generation, international. Columns 2 and 4 include 309 individual fixed effects. Standard errors are clustered by individuals.

the violation is not large enough to overturn our reduced form findings for rent.

Panel B reports an analogous pattern of switching behavior for network shares. Students tended to switch in neighborhoods with high network shares (7.18 percent) and switch out neighborhoods with low network shares (5.72 percent). Always top 3 neighborhoods have the highest average network shares (7.73 percent) and never top 3 neighborhoods have the lowest average network shares (3.87 percent).

Table 4 shows that this pattern is robust to a regression analysis controlling for MSA fixed effects and demographic controls (odd columns) and even individual fixed effects (even columns). In the most saturated specification with individual fixed effects, neighborhoods that are switched in have $432 lower monthly rent and network shares that are greater by 1.46 percentage points on average, relative to neighborhoods that were switched out of the top 3. Our findings are similar if we used top one, top five, or top ten. As described in section 2.1, absent these switchers, pre information rankings would likely be highly collinear with post information rankings, and preferences for rents and network would not be identified.

Our neighborhood choice model with imperfect information will make use of all rankings, not just in and out of top 3. A large number of students (76%) had at least one change in their neighborhood rankings after receiving the information, including 64% who had at least one change in their top 3 neighborhoods. About 50% of the individual neighborhood choices involve a change of at least 2 ranks. Our model also captures remarkable heterogeneity across the individual-level rankings, with close to half of the neighborhoods ranked as a top neighborhood by someone.

Changes in Knowledge of Walk Scores. These changes in rankings provide a useful source of identifying variation in our neighborhood choice model, under the assumption that they are driven by shifts in knowledge about rents and networks. This could fail if individuals also updated their knowledge about other amenities. To assess this, we utilize the survey which asked about walk scores before and after we provided information about rents and network shares.

First,it is encouraging that knowledge of the walk scores remained similar before and after information. A lion's share (81%) of the observations were identical. An additional 10% reflect a decline in the walk score by only one level (out of 4) and an 8% corresponded to an increase in the walk score of one.

More importantly, we show in Table 5 that changes in the walk score are uncorrelated with rents, network shares, and pre information rankings. Column 1 regresses the change in the walk score reported by individual i for neighborhood j (post - pre) on the rents and network shares in neighborhoods, controlling for 308 individual fixed effects and clustering standard errors by individual. Column 2 additionally controls for pre information rankings. All three factors remain uncorrelated with changes in the walk score.

4.3 Neighborhood Preferences Persist After Neighborhood Survey

The stated preference behaviors uncovered in our analysis of the neighborhood choice program are also present in revealed preference behavior during the search and actual neighborhood choices after graduation. Starting with search, Panels A and B of Table 6 present cross-tabulations to compare rents and network shares for neighborhoods ranked in the top 3 pre information, and top 3 in the post-survey search. Our measure of search is based on

Table 5: Walk Score Before and After Information

Dependent variable:	**WalkScore$_{\mathbf{Post}}$ - WalkScore$_{\mathbf{Pre}}$**	
	(1)	(2)
Rent	-0.009	-0.007
	(0.021)	(0.021)
Network	0.005	0.005
	(0.004)	(0.004)
Pre-Ranking		0.007
		(0.007)
N	1295	1295
R-squared	0.32	0.32
Individual FE	Y	Y

* 0.10 ** 0.05 *** 0.01
Notes: This regression uses survey data from 2022 where we asked students to indicate the walk scores for their considered neighborhoods before and after information. The sample includes 1,295 neighborhood-by-individual walk scores (categorized in 4 levels where higher is more walkable). The dependent variable is the post walk score minus the pre walk score. The key regressors include Zillow rents and network shares (column 1), as well as pre information rankings (column 2). All columns control for 308 individual fixed effects. Standard errors are clustered by individuals.

neighborhood clicks from the interactive maps made available to students upon the survey completion, i.e., for each individual we count and rank neighborhoods based on the number of clicks.[15] As with Table 3, we compare actual rents and neighborhood shares for the four groups of neighborhoods.

The diagonal entries show that always top 3 and never top 3 neighborhoods have patterns consistent with stated preferences, i.e., students tend to search more in neighborhoods that are more expensive and with a larger network. More interestingly, we also find consistent behavior for switchers, i.e., students tend to switch into neighborhoods with lower rents and larger networks, while switching out of neighborhoods with higher rents and smaller networks.

Turning to post graduation choices, Panels A and B of Table 7 present cross tabulations for the actual post survey neighborhood choice and the pre intervention top 1 neighborhood. We restrict to top 1 rank because individuals only chose one actual neighborhood to live. Again, the patterns echo those in Table 3, with students having persistent tastes, generally

[15] These maps show a more disagregate level of neighborhood than what is presented in the survey. We aggregate map clicks at the survey level neighborhood before presenting the results. This analysis also uses a more limited number of students since only 44% of individuals used the interactive maps.

Table 6: Rent and Same-School Network for Top Three Neighborhoods Pre and Post Information (Map Clicks)

Panel A: Rent

		Pre	
		Yes	No
Post	Yes	Always Top 3 $3,437	Switch In $3,179
	No	Switch Out $3,672	Never Top 3 $2,931

Panel B: Same-School Network

		Pre	
		Yes	No
Post	Yes	Always Top 3 9.06%	Switch In 7.73%
	No	Switch Out 6.48%	Never Top 3 3.88%

Notes: Repeats Table 3 but using the number of map clicks in each neighborhood to define top three neighborhoods post information.

choosing neighborhoods with higher rents and network shares. Neighborhoods that are always top 1 in the survey and post graduation have higher rent on average ($3,361) and larger network shares (8.12%) relative to never-top-1 neighborhoods which tend to have lower average rent ($3,098) and lower average network shares (4.16%). Once again, switchers influenced by the intervention tend to switch in neighborhoods with lower rents ($3,391) and larger network shares (7.26%) and switch out neighborhoods with higher rent ($3,923) and lower network shares (6.56%). Overall, this translated into savings in rental cost with students switching in neighborhoods where the monthly Zillow rent is $532 lower relative their pre information top-ranked neighborhood stated in their survey (difference of $3,391 and $3,923). Network amenities are similar (difference of 0.7 percentage points). These results indicate that the new neighborhood choice program successfully provided consistent and systematic information that allowed students to improve their location decisions

Table 7: Comparing Neighborhood Choices Post Graduation with Survey Data

Panel A: Rent

		Pre: Yes	Pre: No
Post	Yes	Always Selected $3,361	Switch In $3,391
	No	Switch Out $3,923	Never Selected $3,098

Panel B: Same-School Network

		Pre: Yes	Pre: No
Post	Yes	Always Selected 8.12%	Switch In 7.26%
	No	Switch Out 6.56%	Never Selected 4.16%

Notes: Repeats Table 3 but the pre rankings use top neighborhoods instead of top three and the post rankings use data from actual home addresses after graduation.

by choosing places with lower cost of living while preserving similar network amenities.

Table A4 in the Appendix demonstrates that our conclusions remain similar if we compare location decisions for graduates who participated in the neighborhood choice program to those in the same cohort who did not. Column 1 shows that graduates who participated in the program live in neighborhoods where the monthly rents are lower by $196. Column 2 shows that this is driven by switchers who changed rankings in the survey (rents are significantly lower by $370 for switchers and insignificant for non-switchers). The patterns are similar for network shares.

5 Estimation

This section describes how we integrate the survey design with our generalized neighborhood choice model from section 2 to identify preferences for amenities under imperfect information. We estimate a version of equation 5 in two stages. In the first stage (equation 9), we estimate heterogeneous parameters β_{id} and a set of neighborhood fixed effects δ_{jm}.

The key neighborhood amenities we observe include post information network share A, and post information monthly rent P.[16] The preference heterogeneity terms β_{id} for each of the neighborhood amenities follow equation 2 and are a function of five observed demographic variables: age, gender, married and/or with children, first-generation or minority status, and citizenship status.

$$u_{ijm} = \delta_{jm} + P_{jm}\beta_{id}^{P} + A_{jm}\beta_{id}^{A} + X_{jm}\beta_{id}^{X} + \xi_{ijm} + \varepsilon_{ijm} = V_{ijm} + \varepsilon_{ijm} \tag{9}$$

In the second stage (equation 10), we decompose the neighborhood fixed effects δ_{jm} to recover mean preferences β_o, and also the metropolitan area effects, μ_m. The unobserved errors in equation 10 include the unobserved average neighborhood quality ξ_{jm} and η_{jm} is an idiosyncratic error term:[17]

$$\delta_{jm} = \mu_m + P_{jm}\beta_{o}^{P} + A_{jm}\beta_{o}^{A} + X_{jm}\beta_{o}^{X} + \xi_{jm} + \eta_{jm} \tag{10}$$

5.1 Constructing Latent Quality Indices

So far the neighborhood choice model above has solely used the post information data on rankings, network shares, and rents. Now, we introduce the pre information rankings to construct latent quality indices $g_1(\widetilde{\xi}_{ijm})$ and $g_2(\widetilde{\xi}_{jm})$ to respectively control for unobserved heterogeneity associated with ξ_{ijm} and ξ_{jm}.

To construct $g_1(\widetilde{\xi}_{ijm})$, we use six categorical variables based on whether each individual pre ranked a neighborhood in her consideration set as top 1, 2, 3, 4, 5, or from 6 to 10. We omit from the estimation a dummy for neighborhoods never ranked in the pre information data. Additionally, we include the average of these pre rank dummies and interact these averages with demographics. We include these heterogeneity terms in equation 9. For the second stage, we construct $g_2(\widetilde{\xi}_{jm})$ using six averages of the rank dummies in equation 10. In the results section we report robustness tests that investigate the use of different variables to construct latent quality indices.

[16] We also include a set of 2010 Census characteristics X, including average income, the share of college graduates, and the non-White share.

[17] Since the estimation is in two stages, we decompose ξ_{ijm} in equation 5 into the neighborhood average ξ_{jm} and deviations from the neighborhood mean $\Delta\xi_{ijm}$. The average will be absorbed into the δ fixed effect and will need to be accounted for in the second stage estimation. The deviation from the mean is the key omitted variable in the first stage. To keep the notation simple, from here onwards, we denote $\Delta\xi_{ijm}$ as ξ_{ijm} in equation 9.

To identify MWTP for A, we require that $g_1(\widetilde{\xi}_{ijm})$ and $g_2(\widetilde{\xi}_{jm})$ satisfy the index sufficiency assumption (equation 6). Given the survey design, it is plausible that pre information rankings remain relevant proxies for ξ_{ijm} and ξ_{jm} since preferences for other amenities likely remained stable during the two minutes between rankings. Moreover, we assume that individuals fully updated their knowledge for network shares and cost of living ($\Delta P^1_{ij} = 0$ and $\Delta A^1_{ij} = 0$), the excludability assumption in equation 8. This is likely the case for network shares since the students are told that it was constructed using administrative data. For cost of living, we need that the remaining heterogeneity in ΔP_{ij} is orthogonal to ξ, conditional on our controls. We provide empirical support in a robustness test below (Table 9), showing that conditional on pre rankings, further controlling for ΔP and ΔA will not affect MWTP. The key source of identification is the changes in rankings due to the switchers described in the previous section.

One potential concern of the identification strategy is that the new information about network shares and rents may indirectly cause updates on individual's knowledge about other unobserved amenities. However, we can rule out that such indirect updates are significant enough to overturn the direct effects, as shown in the analysis of switchers in the previous section. We find a number of students who are switchers, and show that switchers are not signficantly different, based on observables, from students who did not switch neighborhoods (Table A3). More importantly, we showed in the previous section that individual views about a third amenity, the walk score, did not significantly change after providing information about rents and networks. Below we will show that including in the model such post-information knowledge about a third amenity does not significantly impact estimates for rents and networks.

5.2 Recovering MWTP

We estimate equation 9 using a rank-ordered logit model (Beggs, Cardell, and Hausman, 1981) based on the post information neighborhood rankings in each individual's consideration set. For each individual i choosing metropolitan area m, our data reveals:

$$U_{ir_{i1}} > U_{ir_{i2}} > \ldots > U_{ir_{iL_i}} \tag{11}$$

where r_{il} denotes the neighborhood that received post information ranking l by individual i. Each individual could rank between two and ten neighborhoods, and we denote the last ranked neighborhood for each individual as L_i. For each market m, we normalize by setting to zero the utility for neighborhoods that are only chosen once. Given the extreme value assumption for ε_{ijm}, the probability of individual i choosing a ranking r_i is:

$$\pi_{ir_i} = P[U_{ir_{i1}} > U_{ir_{i2}} > ... > U_{ir_{iL_i}}] \tag{12}$$

$$= \prod_{l=1}^{L_i-1} \frac{\exp(V_{ir_{il}})}{\sum_{h=l}^{L_i} \exp(V_{ir_{ih}})} \tag{13}$$

We rely on maximum likelihood to estimate the model. The log likelihood function is just the sum of the log of the individual probabilities across all individuals:

$$\mathscr{L} = \sum_{i=1}^{I} \log \pi_{ir_i} \tag{14}$$

This first step returns the heterogeneity in preference parameters and the neighborhood fixed effects that maximize the log likelihood function above, i.e., maximize the probability that each individual makes the correct rank ordering of neighborhoods. We estimate the second step using OLS.

Mean and heterogeneity in willingness-to-pay. We define the average marginal willingness to pay for the same-school network neighborhood characteristic as:

$$MWTP = -\frac{\beta_o^A}{\beta_o^P} \tag{15}$$

whose components are estimated in the second stage decomposition of the estimated neighborhood fixed effects δ_j. Moreover, we combine first and second stage estimations to calculate heterogeneity in marginal willingness to pay for N according to the following formula:

$$MWTP_i = -\frac{\beta_o^A + \sum z_{id}\beta_d^A}{\beta_o^P + \sum z_{id}\beta_d^P} \tag{16}$$

Subsequently we compare this MWTP for a baseline individual, representing the majority groups in all dimensions of heterogeneity, against another individual who shares similar demographics with the exception of just one feature d. Standard errors for both average and heterogeneity in willingness to pay are calculated used the Delta method. For the heterogeneity preference terms in equation 9, the identifying assumption is that the six pre information rank dummies and the average rankings interacted with demographics sufficiently address endogeneity arising from ξ_{ij}. We also assume that heterogeneity in preferences is solely a function of our observed demographic variables, following Bayer, Ferreira, and McMillan (2007).

6 Results

The first column of Table 8 reports estimates of mean preferences (β_o) for the same-school network and for rents, using a version of equation 9 that does not include the latent quality indices. Students have negative preference for higher rents, but the estimate is not statistically different from zero. On the other hand, the network amenity estimate is positive and statistically significant. Those two estimates are then converted into a mean MWTP for the network amenity, according to equation 15, and presented in the third row. Students, on average, are willing to pay \$400 per month in rent (given an average rent of \$3,600 in our estimation sample) to live in a neighborhood with a one percentage point higher network share (the average network share is 6%). Such mean MWTP is likely upward biased because the rent estimate is downward biased.

The second column of Table 8 reports similar estimates, but now for a model that includes the latent quality indices. Interestingly, the rent estimate becomes more negative (-0.55) and statistically different from zero. Moreover, conditional on rent, the network estimate is quite similar to the model in column 1, perhaps because that information was truly new for most students. The combination of those two estimates lead to a new mean MWTP for the network amenity of \$123 per month and is now statistically significant at the 5% level.[18] This estimate is 70% smaller than the comparable mean MWTP without the latent index.

It is reassuring that MWTP estimates change significantly with and without $g_1(\widetilde{\xi}_{ij})$ and $g_2(\widetilde{\xi}_j)$. Importantly, the coefficient on rent becomes more negative once we control for latent quality. We would not see this pattern if individuals completely updated their neighborhood rankings after the information intervention in a way that renders the pre information rankings uninformative. Table A5 presents the key coefficients for our latent quality indices. In line with the relevance condition, column 1 shows that neighborhoods

[18] As a robustness check, we also find that this MWTP estimate remains significant at the 5% level if we clustered standard errors by MSA's (the standard error for the MWTP is 45 instead of 40). Given that we only have 18 MSA's, which is less than the rule-of-thumb of 40 clusters (Angrist and Pischke, 2008), we opted to not cluster standard errors in our primary estimation (p.238).

Table 8: MWTP Estimates With and Without Latent Quality Indices

	(1)	(2)
Same-School Network	0.330***	0.310**
	(0.075)	(0.092)
Rent	-0.180	-0.550***
	(0.118)	(0.126)
Implied MWTP	**$400**	**$123****
	(257)	(40)
MSA FE	Y	Y
Census Characteristics	Y	Y
Latent Quality Indices	N	Y

Notes: Mean preference estimates of the neighborhood choice model using neighborhood-by-individual choices post information intervention. The first stage involves a rank-ordered Logit model using post information neighborhood rankings, neighborhood fixed effects, neighborhood amenities (rent, same-school network shares, as well as Census characteristics - average income, share college graduates, average non-White share). The second stage represents a decomposition of the neighborhood fixed effects from the first stage, including MSA fixed effects. Column 2 adds the latent quality indices, $g_1(\widetilde{\xi}_{ij})$ and $g_2(\widetilde{\xi}_j)$. Table A5 presents the coefficients on the different latent quality indices. Standard errors calculated using the Delta method.

ranked favorably pre information are more likely to be ranked favorably post information as well.

Next, Table 9 shows how we can learn about the nature of the imperfect information bias by constructing different latent quality indices. Column 1 is our baseline estimate and columns 2 to 4 repeat the same model but including different latent quality indices. In column 2, the MWTP changes by 54% to $184 if we only include $g_1(\widetilde{\xi}_{ij})$ in the first stage (equation 9), using the six rank dummies as well as the average rankings interacted with demographics. Column 3 shows the MWTP estimate changes by 61% from $400 to $157 using $g_2(\widetilde{\xi}_j)$ in the delta regression alone. It is useful to know that just adding the six average rankings at the neighborhood level can account for the majority of the bias, even more so than using only individual pre ranking dummies in the first stage.

Pre information knowledge of amenities. In column 4, we augment our baseline model in column 1 with pre information estimates of rents and network shares. In principle, these continuous measures of individual perceptions could also proxy for latent quality to the extent that individuals perceive that higher quality neighborhoods have a larger network

Table 9: Robustness Results for Latent Quality Indices

	(1)	(2)	(3)	(4)	(5)
Same-School Network	0.310**	0.328***	0.293***	0.311**	0.296***
	(0.092)	(0.093)	(0.071)	(0.094)	(0.094)
Rent	-0.550***	-0.389**	-0.408***	-0.552***	-0.547***
	(0.126)	(0.121)	(0.120)	(0.127)	(0.133)
Walk Score					-0.006
					(0.091)
Implied MWTP	**$123****	**$184****	**$157****	**$123****	**$118*****
	(40)	(65)	(54)	(41)	(39)
$g_1(\widetilde{\xi}_{ij})$	Y	Y	N	Y	Y
$g_2(\widetilde{\xi}_j)$	Y	N	Y	Y	Y
Pre rent, Pre network	N	N	N	Y	N

Notes: Column 1 reports our baseline model (column 2 in Table 8). Column 2 only includes $g_1(\widetilde{\xi}_{ij})$ six pre information rank dummies and the average of the rankings interacted with demographics in the first stage. Column 3 uses the averages of the six pre information rankings $g_2(\widetilde{\xi}_j)$. Column 4 adds pre information estimates of same-school network shares and rent to the baseline model. Column 5 uses the average post information neighborhood walk score estimate collected in the 2022 wave. Standard errors calculated using the Delta method.

and higher rent. Interestingly, the MWTP estimate remains stable at $123 suggesting these additional proxies are not adding more information above and beyond our baseline model. Moreover, column 4 also provides empirical support to our assumption that individuals fully updated their knowledge for network shares and cost of living so that prior knowledge is irrelevant. It is reassuring that conditional on pre rankings, MWTP estimates are similar with and without controlling for pre-information P and A.

Post information knowledge of amenities. Finally, in column 5, we add a proxy for post information knowledge of amenities. Specifically, we average the post information walk scores in the 2022 survey to the neighborhood level. We then interact this neighborhood characteristic with individual demographics in the first stage, and control for it in the second stage delta regression.

Further controlling for the walk score leads to small changes in our structural estimates. The walk score has a small and insignificant impact of -0.006. The coefficient on rent is also very stable (-0.547 versus -0.550) and the coefficient on networks is slightly smaller

(0.296 versus 0.310). The MWTP is slightly lower ($118 relative to $123).[19]

Heterogeneity. Table 10 reports the implied heterogeneity in MWTP for the network amenity. Panel A presents the overall MWTP across the different specifications and Panel B presents the heterogeneous WTP estimates. There is some heterogeneity in MWTP estimates, but the results are understandably less precise given our sample size. Four measures of heterogeneity generally have negative values relative to the baseline: Age, married and/or with children, international, first-generation and under-represented minorities. Estimates for female are positive and small.

Columns 2 and 3 of Table 10 report results that additionally include measures of student heterogeneity by type of industry and whether the student previously worked in that same MSA. The mean MWTP remains practically unchanged in all models, and none of the extra heterogeneity measures are statistically different from zero. Next we include a variable from the survey related to the stage of the housing search process, i.e., whether the student had already visited rental properties or signed a lease. Accounting for the stage of the search process does not seem to change our main MWTP estimate. Column 5 reports heterogeneity by major. Finally, Column 6 includes all measures of heterogeneity in one model. The mean MWTP from this model is $104.

How do students interpret the new information about the share of same-school network living in a neighborhood? At the very end of the survey we asked students to consider a neighborhood with a large network, and asked them to check at most three reasons for why that neighborhood would be desirable. The most common response was related to professional networking opportunities and social activities with the same-school network. Next came good restaurants and shops, which is not only an indicator of local availability of services, but also an indicator of venues that promote social interactions within the network. Convenient commute was next, followed by high income and well-educated neighbors, and safe area with good schools and parks.

7 Conclusion

We introduce a generalized neighborhood choice model to estimate preferences for amenities under imperfect information. This is a complex and pervasive problem given the many

[19] In Appendix Table A6 we find qualitatively similar patterns for MWTP estimates when using the 2022 survey data with and without walk scores. See Ferreira and Wong (2022) for a discussion of changes in MWTP after the COVID pandemic.

Table 10: Robustness Results for Demographic Attributes

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Overall MWTP						
Implied MWTP	**$123*** **	**$110*** **	**$130*** **	**$103*** **	**$114*** **	**$104*** **
	(40)	(37)	(38)	(37)	(38)	(29)
Panel B: Heterogeneity in MWTP						
Age	-12	-20	-14	4	-20	-25
	(45)	(49)	(31)	(36)	(45)	(31)
Married/Kids	-1,031	-992	-3,300	808	-1,466	309
	(4,400)	(5,284)	(55,334)	(6,354)	(9,055)	(1,299)
International	-83	-68	-14	-55	-93	12
	(108)	(102)	(89)	(94)	(107)	(79)
First-gen/URM	-428	-433	-249*	-317	-427	-194*
	(413)	(445)	(147)	(243)	(435)	(116)
Female	49	52	55	73	40	56
	(82)	(78)	(60)	(71)	(84)	(55)
Industry - Consulting		-50				-57
		(94)				(62)
Industry - Finance		61				47
		(130)				(76)
Previous worked			128			112
			(192)			(154)
Visited/Signed				54		62
				(92)		(68)
Major - Finance/Real estate					-20	-42
					(98)	(63)
Demographics	Y	Y	Y	Y	Y	Y
Industry	N	Y	N	N	N	Y
Previous worked	N	N	Y	N	N	Y
Visited/Signed	N	N	N	Y	N	Y
Major	N	N	N	N	Y	Y

Notes: This table presents robustness estimates with additional observed heterogeneity. Panel A reports the overall MWTP while Panel B reports the heterogeneous WTP estimates. Heterogeneity in WTP estimates are relative to a representative students with median age, single, male, national, and not minority. Column 1 repeats the baseline (column 2 in Table 8). Column 2 adds industry dummies, column 3 adds a dummy for individuals who previously worked in the MSA, column 4 adds a dummy for individuals who have visited or signed a lease, column 5 controls for intended majors. Column 6 includes all controls.

amenities and the large number of neighborhoods in a labor market. Moreover, we show that the endogeneity problem cannot be solved with standard methods since the nature of the imperfect information is heterogeneous across individuals and has a high dimension. To address this, we introduce latent quality indices which we estimate by observing the same individual choosing neighborhoods before and after receiving information about amenities.

Our empirical strategy integrates structural estimation with a survey design associated with a neighborhood choice program for graduating students of a large professional school. The program provided information about cost of living and same-school professional and social network. We observe switchers changing their neighborhood rankings after the information intervention to increase network shares by 1.46 percentage points and decrease rents by $430 for their top-three neighborhoods, implying a positive willingness-to-pay for the same-school network. Similar results are found in revelead preference data from actual neighborhood choices.

The structural preference estimates show that controlling for the latent quality index significantly reduces bias in the marginal utility for rents, implying a mean MWTP for network shares of $123 (with the latent quality index) relative to $400 (without the quality index). We probe the robustness by examining additional demographic variables and different ways to estimate the latent quality index.

Finally, the empirical framework developed in this paper can be fruitfully applied to estimate preferences for any other neighborhood amenity. It can also be applied in many other settings where imperfect information is pervasive or in which individuals have difficulties processing information about choice sets and product characteristics. Examples faced by young adults include choice of college, college major, and type and location (city) of first job.

References

Allcott, H., R. Diamond, J.-P. Dubé, J. Handbury, I. Rahkovsky, and M. Schnell (2019, 05). Food Deserts and the Causes of Nutritional Inequality. *The Quarterly Journal of Economics 134*(4), 1793–1844.

Almagro, M. and T. Domínguez-Iino (2020). Location Sorting and Endogenous Amenities: Evidence from Amsterdam.

Angrist, J. D. and J.-S. Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Bailey, M., R. Cao, T. Kuchler, J. Stroebel, and A. Wong (2018). Social Connectedness: Measurement, Determinants, and Effects. *Journal of Economic Perspectives 32*(3), 259–80.

Bayer, P., F. Ferreira, and R. McMillan (2007). A Unified Framework for Measuring Preferences for Schools and Neighborhoods. *Journal of Political Economy 115*(4), 588–638.

Bayer, P., N. Keohane, and C. Timmins (2009). Migration and Hedonic Valuation: The Case of Air Quality. *Journal of Environmental Economics and Management 58*(1), 1–14.

Bayer, P., R. McMillan, A. Murphy, and C. Timmins (2016). A Dynamic Model of Demand for Houses and Neighborhoods. *Econometrica 84*(3), 893–942.

Bayer, P., S. Ross, and G. Topa (2008). Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes. *Journal of Political Economy 116*(6), 1150–1196.

Beggs, S., S. Cardell, and J. Hausman (1981). Assessing the Potential Demand for Electric Cars. *Journal of Econometrics 17*(1), 1–19.

Benjamin, D. J., O. Heffetz, M. S. Kimball, and A. Rees-Jones (2014, November). Can Marginal Rates of Substitution Be Inferred from Happiness Data? Evidence from Residency Choices. *American Economic Review 104*(11), 3498–3528.

Bergman, P., E. W. Chan, and A. Kapor (2020). Housing Search Frictions: Evidence from Detailed Search Data and a Field Experiment. Working Paper 27209, National Bureau of Economic Research.

Bergman, P., R. Chetty, S. DeLuca, N. Hendren, L. Katz, and C. Palmer (2020). Creating Moves to Opportunity: Experimental Evidence on Barriers to Neighborhood Choice. Working paper.

Berry, S. T. (1994). Estimating Discrete-Choice Models of Product Differentiation. *The Rand Journal of Economics 25*(2), 242–262.

Berry, S. T. and P. A. Haile (2014). Identification in Differentiated Products Markets Using Market Level Data. *Econometrica 82*(5), 1749–1797.

Berry, S. T., J. Levinsohn, and A. Pakes (1995). Automobile Prices in Market Equilibrium. *Econometrica 63*(4), 841–890.

Berry, S. T., J. Levinsohn, and A. 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.

Bottan, N. L. and R. Perez-Truglia (2022). Choosing Your Pond: Location Choices and Relative Income. *The Review of Economics and Statistics*.

Büchel, K., M. V. Ehrlich, D. Puga, and E. Viladecans-Marsal (2020). Calling from the Outside: The Role of Networks in Residential Mobility. *Journal of Urban Economics 119*, 103277.

Caetano, G. and V. Maheshri (2019). A Unified Empirical Framework to Study Segregation. Working paper.

Calder-Wang, S. (2019). The Distributional Impact of the Sharing Economy on the Housing Market. Working paper.

Chay, K. Y. and M. Greenstone (2003, 08). The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession. *The Quarterly Journal of Economics 118*(3), 1121–1167.

Chay, K. Y. and M. Greenstone (2005). Does Air Quality Matter? Evidence from the Housing Market. *Journal of Political Economy 113*(2), 376–424.

Chetty, R., J. N. Friedman, N. Hendren, M. R. Jones, and S. R. Porter (2018, October). The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility. Working Paper 25147, National Bureau of Economic Research.

Chetty, R. and N. Hendren (2018a, 02). The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects. *The Quarterly Journal of Economics 133*(3), 1107–1162.

Chetty, R. and N. Hendren (2018b, 02). The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates. *The Quarterly Journal of Economics 133*(3), 1163–1228.

Chetty, R., N. Hendren, and L. F. Katz (2016, April). The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment. *American Economic Review 106*(4), 855–902.

Dahl, G. B. (2002). Mobility and the Return to Education: Testing a Roy Model with Multiple Markets. *Econometrica 70*, 2367–2420.

Diamond, R. (2016). The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000. *American Economic Review 106*(3), 479–524.

Diamond, R., T. McQuade, and F. Qian (2019, September). The effects of rent control expansion on tenants, landlords, and inequality: Evidence from san francisco. *American Economic Review 109*(9), 3365–94.

Epple, D., L. Quintero, and H. Sieg (2020). A New Approach to Estimating Equilibrium Models for Metropolitan Housing Markets. *Journal of Political Economy 128*(3), 948–983.

Epple, D. and H. Sieg (1999). Estimating Equilibrium Models of Local Jurisdictions. *Journal of Political Economy 107*(4), 645–681.

Ferreira, F. and M. Wong (2022). Neighborhood Choice After COVID: The Role of Rents, Amenities, and Work-From-Home. Working Paper 29960, National Bureau of Economic Research.

Fujiwara, T., E. Morales, and C. Porcher (2019). A Revealed-Preference Approach to Measuring Information Frictions in Migration. Working paper.

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

Gao, Z., M. Sockin, and W. Xiong (2021). Learning about the Neighborhood. *The Review of Financial Studies 34*(9), 4323–4372.

Honoré, B. E. and L. Hu (2020). Selection without Exclusion. *Econometrica 88*(3), 1007–1029.

Isen, A., M. Rossin-Slater, and W. R. Walker (2017). Every Breath You Take – Every Dollar You'll Make: The Long-Term Consequences of the Clean Air Act of 1970. *Journal of Political Economy 125*(3), 848–902.

Kacperczyk, M., S. Van Nieuwerburgh, and L. Veldkamp (2016). A Rational Theory of Mutual Funds' Attention Allocation. *Econometrica 84*(2), 571–626.

Kling, J., J. Liebman, and L. F. Katz (2007). Experimental Analysis of Neighborhood Effects. *Econometrica 75*(1), 83–119.

Kosar, G., T. Ransom, and W. Van der Klaauw (2021). Understanding Migration Aversion Using Elicited Counterfactual Choice Probabilities. *Journal of Econometrics*.

Linden, L. and J. E. Rockoff (2008, June). Estimates of the Impact of Crime Risk on Property Values from Megan's Laws. *American Economic Review 98*(3), 1103–27.

Manski, C. F. (1988). Identification of Binary Response Models. *Journal of the American Statistical Association 83*(403), 729–738.

McFadden, D. (1973). Conditional Logit Analysis of Qualitative Choice Behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics*. New York: Academy Press.

McFadden, D. (1978). Modelling the Choice of Residential Location. In A. Karlqvist (Ed.), *Spatial Interaction Theory and Planning Models*. New York: North-Holland.

Moretti, E. (2013). *The New Geography of Jobs*. Mariner Books.

National Center for Education Statistics (2020). Digest of Education Statistics.

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

Pakes, A. and J. Porter (2016, January). Moment Inequalities for Multinomial Choice with Fixed Effects. Working Paper 21893, National Bureau of Economic Research.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2015). Moment Inequalities and Their Application. *Econometrica 83*(1), 315–334.

Petrin, A. (2002). Quantifying the Benefits of New Products: The Case of the Minivan. *Journal of Political Economy 110*(4), 705–729.

Qian, F. (2021). The Effects of High-skilled Firm Entry on Incumbent Residents.

Shue, K. (2013). Executive Networks and Firm Policies: Evidence from the Random Assignment of MBA Peers. *Review of Financial Studies 26*(6), 1401–1442.

Train, K. (2009). *Discrete Choice Methods With Simulation*. Cambridge, UK: Cambridge University Press.

Wiswall, M. and B. Zafar (2014). Determinants of College Major Choice: Identification using an Information Experiment. *The Review of Economic Studies 82*(2), 791–824.

Wong, M. (2013). Estimating Ethnic Preferences Using Ethnic Housing Quotas in Singapore. *Review of Economic Studies 80*(3), 1178–1214.

Online Appendix Tables (Ferreira and Wong, 2021)

Table A1: Characteristics of Respondents and Full Sampling Frame

	All	Respondents	Non-respondents	Difference	p-value
Panel A: Survey respondents in top 20 MSA's					
Female	0.44	0.50	0.41	0.09***	0.01
Age	29.68	29.36	29.87	-0.51***	0.00
Married/Kids	0.13	0.12	0.14	-0.02	0.33
First-gen/URM	0.19	0.15	0.20	-0.05*	0.06
International	0.34	0.23	0.40	-0.17***	0.00
N	852	309	543	852	852
Panel B: Survey respondents choosing all cities					
Female	0.44	0.49	0.41	0.07**	0.03
Age	29.68	29.46	29.83	-0.37**	0.01
Married/Kids	0.13	0.14	0.13	0.01	0.69
First-gen/URM	0.19	0.16	0.20	-0.05*	0.10
International	0.34	0.28	0.38	-0.09***	0.00
N	852	341	511	852	852

* 0.10 ** 0.05 *** 0.01

Notes: Panels A and B show how the demographics for survey respondents compare to the full student population of 852 students. Panel A includes the 309 students in our primary estimation sample, i.e. those who chose the top MSA's in our program. Panel B includes 341 students who responded to the survey, including 32 who chose other cities not in the Neighborhood Choice Program. The five demographic characteristics include an indicator for females, age, an indicator for married individuals or those who have children, an indicator for first-generation or under-represented minorities, an indicator for international students who are not U.S. citizens.

Table A2: Students in the Map and Post Graduation Samples Relative to Survey Respondents

	Survey	Post Survey	Missing	Difference	p-value
Panel A: Map clicks					
Female	0.50	0.43	0.54	-0.10*	0.09
Age	29.36	29.51	29.28	0.23	0.28
Married/Kids	0.12	0.16	0.10	0.06	0.11
First-gen/URM	0.15	0.17	0.14	0.03	0.53
International	0.23	0.22	0.24	-0.02	0.63
N	309	106	203	309	309
Panel B: Post graduation choices					
Female	0.50	0.51	0.49	0.02	0.69
Age	29.36	29.38	29.32	0.06	0.75
Married/Kids	0.12	0.10	0.15	-0.05	0.15
First-gen/URM	0.15	0.16	0.14	0.03	0.48
International	0.23	0.18	0.31	-0.13***	0.01
N	309	176	133	309	309

* 0.10 ** 0.05 *** 0.01

Notes: Similar to Table A1 but Panel A compares demographics for 106 students in the map clicks data relative to the survey respondents that are missing map clicks. Panel B compares demographics for 176 students we have post graduation location data relative to those with missing addresses.

Table A3: Demographics for Switchers and Non-Switchers

	Always in top 3	Switch in	Switch out	Switch in - Switch out
Female	0.00	0.01	0.01	0.00
	[0.64]	[0.71]	[0.64]	[0.87]
Age	-0.01	0.02	0.02	-0.01
	[0.83]	[0.87]	[0.81]	[0.82]
Married/Kids	0.01	0.00	0.00	-0.01
	[0.37]	[0.98]	[0.74]	[0.36]
First-gen/URM	0.00	0.01	0.01	0.01
	[0.84]	[0.54]	[0.71]	[0.36]
International	-0.03***	0.08***	0.07***	0.01
	[0.00]	[0.00]	[0.00]	[0.45]
N	7012	7012	7012	7012

* 0.10 ** 0.05 *** 0.01

Notes: Each row repeats the OLS regression in column 1 of Table 4 but the dependent variables are now student demographics instead of Zillow rent or network shares. We include MSA fixed effects but no demographic controls. Standard errors clustered by individuals.

Table A4: Rent and Same-School Network for Actual Neighborhood Choices

Dependent variable:	**Rent**		**Same-School Network**	
	(1)	(2)	(3)	(4)
Survey	-0.20**	0.10	-0.30	-1.37*
	(0.09)	(0.15)	(0.44)	(0.78)
Survey*Switcher		-0.37**		1.35*
		(0.16)		(0.80)
N	365	365	365	365
R-squared	0.51	0.52	0.49	0.50
Demographics	Y	Y	Y	Y
MSA FE	Y	Y	Y	Y

* 0.10 ** 0.05 *** 0.01
Notes: OLS regressions comparing post-graduation neighborhood choices for graduates who participated in the program and those in the same cohort who did not. The sample includes 365 students who graduated in 2019 and live in cities in the neighborhood choice program. The sample excludes the city that the school is in as we define post-graduation address if there is a change in location after graduation. We control for MSA fixed effects and demographic characteristics, similar to column 1 in Table 4. Column 2 adds an interaction with an indicator for survey respondents who are switchers (those who changed their rankings for more than half of the neighborhoods considered, 80 percent of program participants in the estimation sample). Results are similar using different cutoffs. Columns 1 and 2 examine rents. Columns 3 and 4 repeat the same for network shares.

Table A5: Estimates for Latent Quality Indices

	(1)	(2)	(3)	(4)	(5)
Same-School Network	0.310**	0.328***	0.293***	0.311**	0.296***
	(0.092)	(0.093)	(0.071)	(0.094)	(0.094)
Rent	-0.550***	-0.389**	-0.408***	-0.552***	-0.547***
	(0.126)	(0.121)	(0.120)	(0.127)	(0.133)
Walk Score					-0.006
					(0.091)
Implied MWTP	**$123****	**$184****	**$157****	**$123****	**118*****
	(40)	(65)	(54)	(41)	(39)
Rank 1	0.711***	0.711***		0.712***	0.723***
	(0.065)	(0.065)		(0.076)	(0.066)
Rank 2	0.436***	0.436***		0.438***	0.436***
	(0.063)	(0.063)		(0.074)	(0.063)
Rank 3	0.205***	0.205***		0.208**	0.213***
	(0.057)	(0.057)		(0.067)	(0.058)
Rank 4	0.033	0.033		0.037	0.036
	(0.055)	(0.055)		(0.064)	(0.055)
Rank 5	-0.061	-0.061		-0.056	-0.058
	(0.052)	(0.052)		(0.059)	(0.052)
Rank 6-plus	-0.209**	-0.209**		-0.202**	-0.201***
	(0.065)	(0.065)		(0.076)	(0.066)
Average Rank 1	0.296***		0.377***	0.298***	0.303***
	(0.076)		(0.070)	(0.077)	(0.082)
Average Rank 2	0.165*		0.233***	0.165*	0.173***
	(0.064)		(0.062)	(0.065)	(0.066)
Average Rank 3	0.086		0.149**	0.089	0.100
	(0.066)		(0.057)	(0.067)	(0.071)
Average Rank 4	0.092		0.200***	0.086	0.085
	(0.065)		(0.045)	(0.068)	(0.072)
Average Rank 5	0.087		0.118*	0.088	0.063
	(0.074)		(0.056)	(0.076)	(0.087)
Average Rank 6-plus	0.138		0.086	0.141	0.109
	(0.096)		(0.073)	(0.097)	(0.108)
Pre information rent				0.006	
				(0.065)	
Pre information network shares				0.067	
				(0.096)	
$g_1(\widetilde{\xi}_{ij})$	Y	Y	N	Y	Y
$g_2(\widetilde{\xi}_j)$	Y	N	Y	Y	Y
Pre rent, Pre network	N	N	N	Y	N

* 0.10 ** 0.05 *** 0.01

Notes: Coefficients on the latent quality indices for each of the four specifications reported in Table 9. In column 1, we present estimates for our baseline model by reporting coefficients for the six individual rank dummies in the first stage $g_1(\widetilde{\xi}_{ij})$ and the six average rank dummies in the second stage ($g_2(\widetilde{\xi}_j)$). We suppress interactions of the average rankings with demographics due to space constraints. Column 2 only includes $g_1(\widetilde{\xi}_{ij})$ and column 3 only includes $g_2(\widetilde{\xi}_j)$. Column 4 includes both and adds the pre information estimate of rent and network shares. Column 5 uses the average post information neighborhood walk score estimate collected in the 2022 wave.

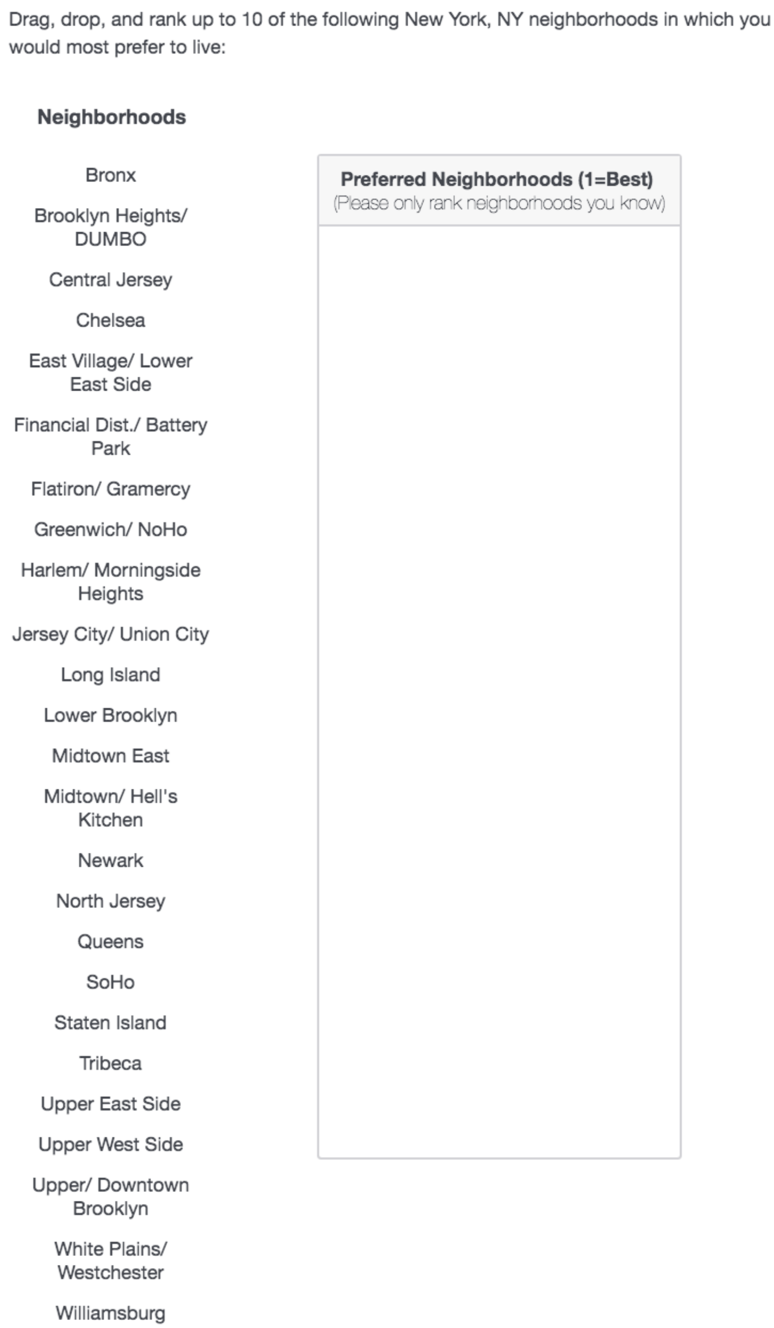
Table A6: MWTP Estimates With and Without Walk Scores Using 2022 Survey

	(1)	(2)
Same-School Network	0.368**	0.309***
	(0.090)	(0.083)
Rent	-0.190	-0.160
	(0.154)	(0.158)
Walk Score		0.181
		(0.168)
Implied MWTP	**$391**	**$391**
	(318)	(383)
MSA FE	Y	Y
Census Characteristics	Y	Y
Latent Quality Indices	Y	Y

Notes: Column 1 reports our baseline model (column 2 in Table 8) using the 2022 survey data. Column 2 adds the average post information neighborhood walk score. Standard errors calculated using the Delta method.

Survey Appendix

Figure A1: Pre Information Choice Set


Drag, drop, and rank up to 10 of the following New York, NY neighborhoods in which you would most prefer to live:

Neighborhoods

Bronx
Brooklyn Heights/ DUMBO
Central Jersey
Chelsea
East Village/ Lower East Side
Financial Dist./ Battery Park
Flatiron/ Gramercy
Greenwich/ NoHo
Harlem/ Morningside Heights
Jersey City/ Union City
Long Island
Lower Brooklyn
Midtown East
Midtown/ Hell's Kitchen
Newark
North Jersey
Queens
SoHo
Staten Island
Tribeca
Upper East Side
Upper West Side
Upper/ Downtown Brooklyn
White Plains/ Westchester
Williamsburg

Preferred Neighborhoods (1=Best)
(Please only rank neighborhoods you know)

Figure A2: Pre Information Ranking of Neighborhoods

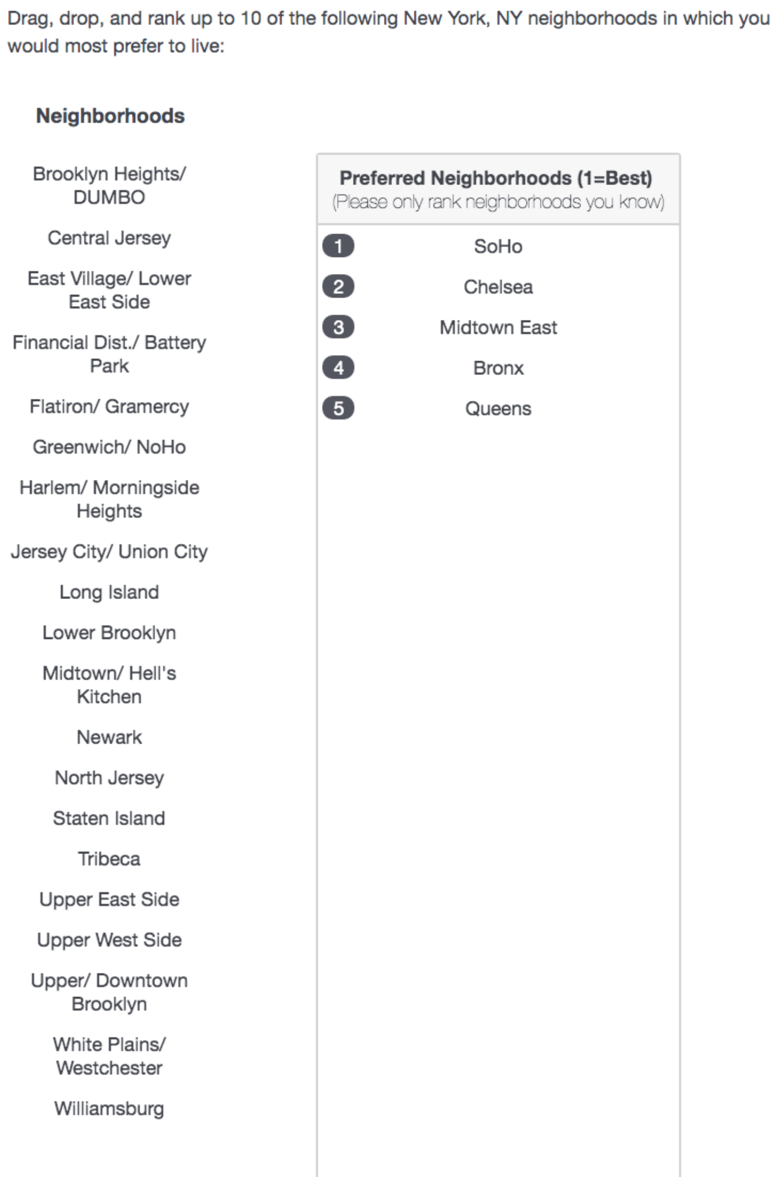


Figure A3: Estimates of Monthly Rent in Considered Neighborhoods

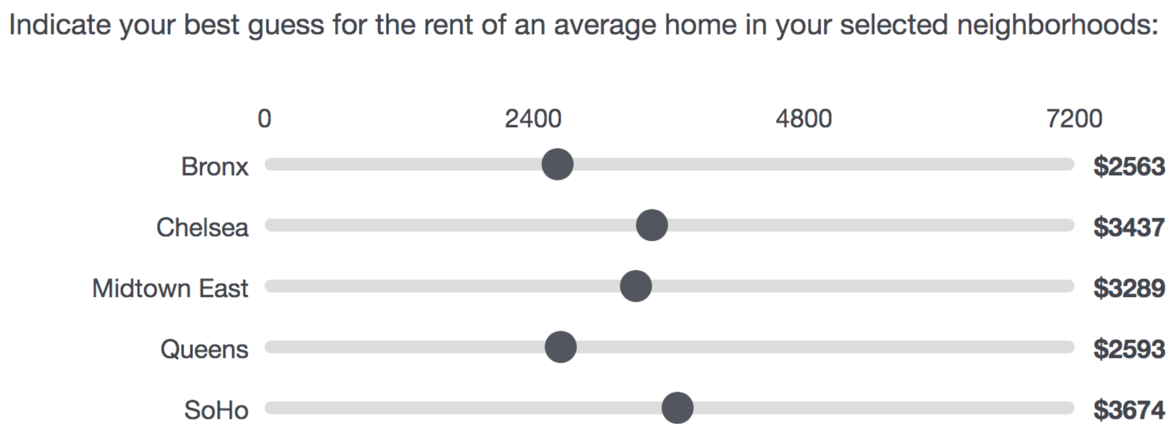


Figure A4: Estimates of Same-School Network Shares in Considered Neighborhoods

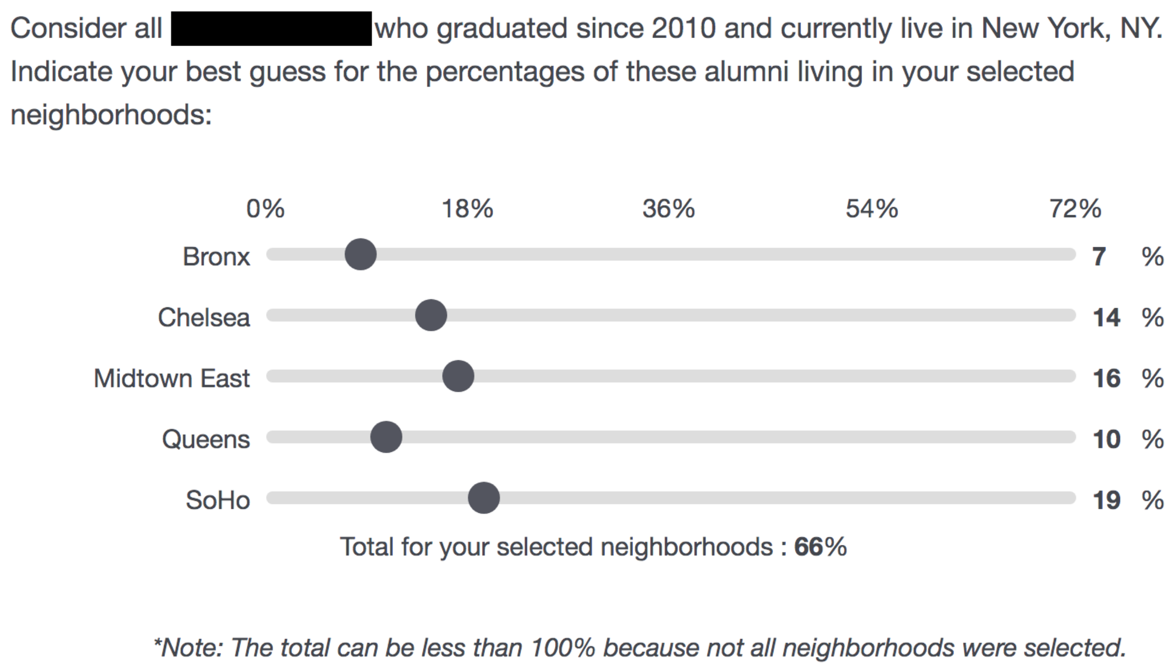
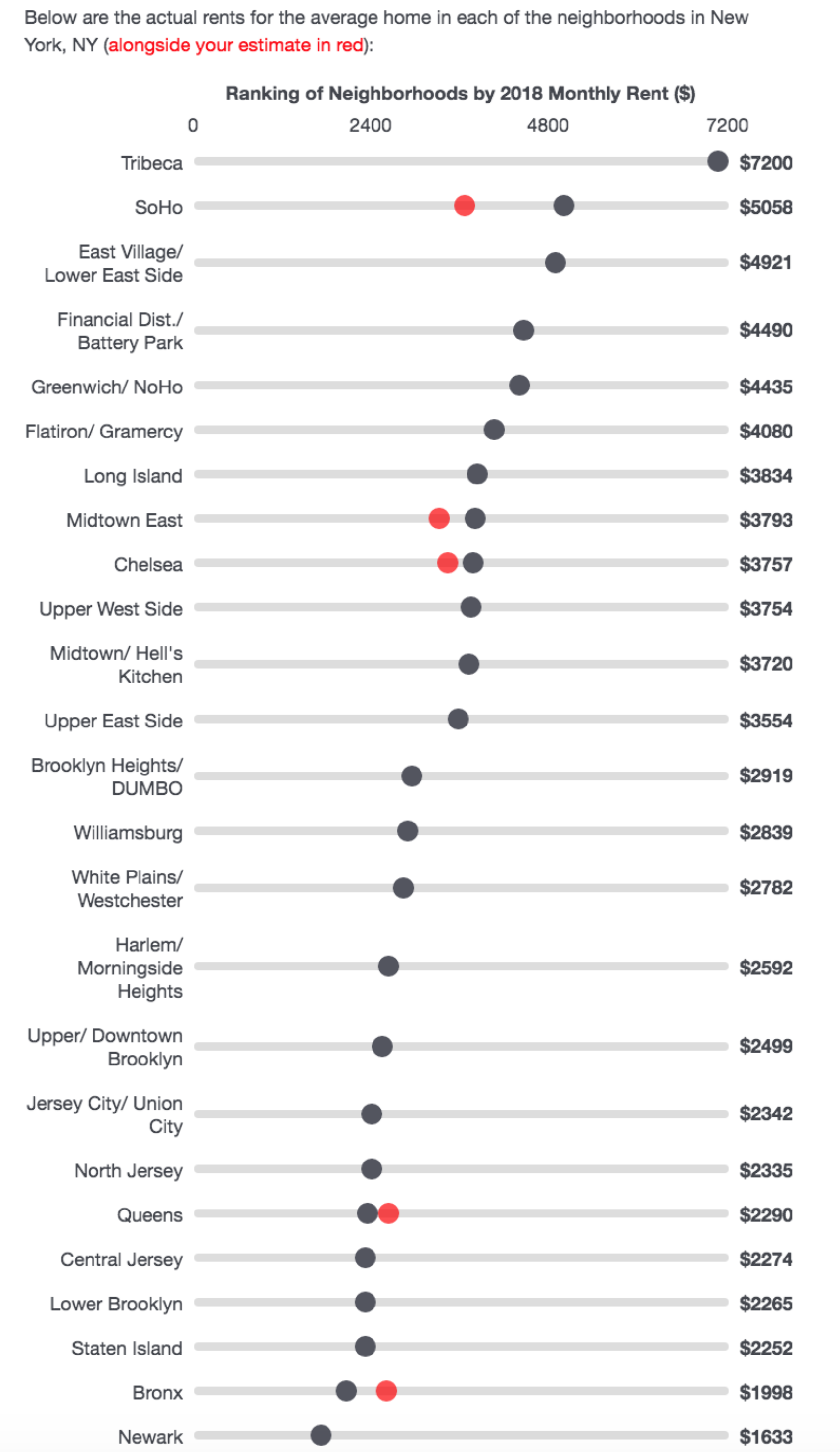


Figure A5: Monthly Zillow Rent (Alongside Pre Information Estimates) in All Neighborhoods



Notes: We also presented an analogous figure for same-school network shares but suppressed it here due to the proprietary nature of the data.

Figure A6: Post Information Ranking of Neighborhoods


Please update your ranking of preferred neighborhoods:

Neighborhoods

Bronx
[Rent: $1998, Alumni: [redacted]%]

Brooklyn Heights/ DUMBO
[Rent: $2919, Alumni: [redacted]%]

Central Jersey
[Rent: $2274, Alumni: [redacted]%]

Chelsea
[Rent: $3757, Alumni: [redacted]%]

East Village/ Lower East Side
[Rent: $4921, Alumni: [redacted]%]

Financial Dist./ Battery Park
[Rent: $4490, Alumni: [redacted]%]

Flatiron/ Gramercy
[Rent: $4080, Alumni: [redacted]%]

Greenwich/ NoHo
[Rent: $4435, Alumni: [redacted]%]

Harlem/ Morningside Heights
[Rent: $2592, Alumni: [redacted]%]

Jersey City/ Union City
[Rent: $2342, Alumni: [redacted]%]

Long Island
[Rent: $3834, Alumni: [redacted]%]

Lower Brooklyn
[Rent: $2265, Alumni: [redacted]%]

Midtown East
[Rent: $3793, Alumni: [redacted]%]

Midtown/ Hell's Kitchen
[Rent: $3720, Alumni: [redacted]%]

Newark
[Rent: $1633, Alumni: [redacted]%]

North Jersey
[Rent: $2335, Alumni: [redacted]%]

Queens
[Rent: $2290, Alumni: [redacted]%]

SoHo
[Rent: $5058, Alumni: [redacted]%]

Staten Island
[Rent: $2252, Alumni: [redacted]%]

Tribeca
[Rent: $7675, Alumni: [redacted]%]

Upper East Side
[Rent: $3554, Alumni: [redacted]%]

Upper West Side
[Rent: $3754, Alumni: [redacted]%]

Upper/ Downtown Brooklyn
[Rent: $2499, Alumni: [redacted]%]

White Plains/ Westchester
[Rent: $2782, Alumni: [redacted]%]

Williamsburg
[Rent: $2839, Alumni: [redacted]%]

Preferred Neighborhoods (1=Best)
(Please only rank neighborhoods you know)

Notes: Survey repondents saw a full schedule of all neighborhoods, as well as the Zillow rent and network shares (suppressed here due to the proprietary nature of the data).