

# The Effects of High-skilled Firm Entry on Incumbent Residents\*

Franklin Qian<sup>†</sup>

Rose Tan<sup>‡</sup>

July 27, 2021

Click [here](#) for the latest version.

## Abstract

What happens to incumbent residents following the entry of a large high-skilled firm? To study this, we construct a dataset of 391 such entries in the U.S. from 1990–2010. We follow incumbent residents over 13 years using rich micro-data on individual address histories, property characteristics, and financial records. First, we estimate the effects of the firm entry on incumbent residents’ consumption, finances, and mobility. To do so, we compare outcomes for residents living close to the entry location with those living far away, while controlling for their proximity to potential high-skilled firm entry sites. Next, we decompose welfare from changes in wages, rents, and amenities for incumbent residents using a model of individual home and work location choice. Taken together, our results show high-skilled incumbents, especially homeowners, benefit. Low-skilled owners benefit less than high-skilled owners. Low-skilled renters are harmed. In the medium to long run, they incur an annual welfare loss that is equivalent to a 0.2 percent decline in their wages one year prior to the entry.

---

\*We thank Nicholas Bloom, Rebecca Diamond, Liran Einav and Matthew Gentzkow for their invaluable guidance. We also thank Luis Armona, Shai Bernstein, Laura Blattner, Alec Brandon, Raj Chetty, Mark Duggan, Benjamin Klopach, Caroline Hoxby, Erik Hurst, Guido Imbens, Timothy McQuade, Scott Nelson, Michael Pollmann, Issac Sorkin, Melanie Wallskog, and many seminar participants for helpful comments and suggestions. Qian acknowledges support by the B.F. Haley and E.S. Shaw Fellowship for Economics and the Shultz Graduate Student Fellowship in Economic Policy through a grant to the Stanford Institute for Economic Policy Research. Tan acknowledges support from the National Science Foundation Graduate Research Fellowship (Grant #1656518). Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Stanford Institute for Economic Policy Research, the National Science Foundation or LinkedIn. The authors have obtained IRB approval from Stanford University to conduct this research.

<sup>†</sup>UNC Kenan-Flagler Business School. Email: [Franklin.Qian@kenan-flagler.unc.edu](mailto:Franklin.Qian@kenan-flagler.unc.edu).

<sup>‡</sup>LinkedIn. Email: [rose2@alumni.stanford.edu](mailto:rose2@alumni.stanford.edu).

# 1 Introduction

In recent decades, local governments in the U.S. have competed fiercely to attract large high-skilled firms. They offer substantial tax subsidies in the hopes that these firms will create jobs and generate economic growth. One salient example is Amazon—218 cities competed for its second headquarters ([Goodman \(2018\)](#)). However, there is an intense debate on whether high-skilled firm entries can benefit not only the local high-skilled workers but also the low-skilled workers. In a 2018 survey conducted among over 200 mayors in the U.S., 84% of mayors think recruiting jobs and investment to their city with financial incentives is good for the city. However, 44% of them think it is unpopular with their constituents ([Einstein et al. \(2018\)](#)). On the one hand, supporters are excited about the new jobs that will generate economic growth. On the other hand, opponents fear gentrification, rising inequality, and the heavy price tags of tax subsidies. Do low-skilled workers benefit when a large firm’s establishment employing mainly high-skilled workers opens in an area? Despite the potentially large distributional consequences, we lack empirical research evidence on this question. This paper fills this gap by quantifying the welfare consequences of high-skilled firm entry on incumbent residents.

Whether low-skilled incumbent residents benefit from high-skilled firm entry is theoretically ambiguous. Economic theory predicts that when a high-skilled firm enters, low-skilled workers may be impacted by rising house prices, changes in wages and changes in amenities. First, rising house prices hurt renters, but may be neutral or beneficial for homeowners. Second, high-skilled wages increase after the entry, but low-skilled wages may or may not increase. Third, amenities may change such that they no longer suit the preferences of low-skilled incumbents; they also change to improve local public goods. The effects of these changes on low-skilled incumbents are ambiguous regardless of whether they stay or move away from their original neighborhoods.

In this paper, we bring to bear novel microdata and use a sample of 391 high-skilled firm entries in the U.S. during 1990–2010 to quantify the welfare incidence of high-skilled firm entry on local residents. We first estimate the effects of high-skilled firm entry on incumbent residents’ consumption, credit events, and mobility that are due to differential exposure to the firm entry shock. After matching on residential neighborhoods based on their proximity to potential high-skilled firm entry locations, we compare incumbent residents initially living close to the true entry location vs. those far away.

Next, we estimate a model of worker’s home and work location choices to quantify the welfare incidence due to a representative high-skilled firm entry. The results from the model are broadly consistent with the reduced-form evidence. We do confirm that incumbent low-skilled renters are harmed. Low-skilled homeowners who are a substantial share of the overall low-skilled incumbents are actually not harmed, or may even benefit. On aggregate, low-skilled incumbents lose by an annual \$13.2 million. On the other hand, we find substantial gains for high-skilled incumbents, especially high-skilled owners. On aggregate, high-skilled incumbents benefit by an annual \$23.2 million, which makes the total welfare impact of a representative high-skilled firm entry clearly positive.

To conduct our analysis, we construct a panel dataset that follows 45 million individuals who initially live within 30 minutes of driving time from each of the 391 entry sites. The firm entry announcements, obtained from Conway Analytics, promised at least 500 new jobs for high-skilled workers and occurred between 1990 and 2010. Using address history records from Infutor, we follow residents in the four years prior and eight years after the announcement of firm entry.

We obtain incumbent residents’ education level and homeownership status, as well as their financial information. To determine homeownership status, we merge names from Infutor with historical housing tax and deed records from CoreLogic, which allows us to identify in each year from 1990 onward whether someone is an owner or renter. To determine education, which proxies for skill group, we build a machine learning model that uses individual characteristics from the ACS to predict education, and then we use the model to predict education out of sample on the Infutor data<sup>1</sup>. We obtain financial information from TransUnion (TU) for an anonymized random subsample of about 8 million individuals. TU provides credit scores, loan balances, and delinquencies for major categories of debt including mortgages and auto loans.

We begin our analysis by examining the effects of high-skilled firm entry on neighborhood characteristics and incumbent residents’ outcomes due to differential exposure to the firm entry shock. Because firms choose their entry locations endogenously, simply comparing neighborhoods close to and far away from the entry site could omit important differences in neighborhoods that are correlated with the likelihood of attracting a firm entry nearby. For example, they could differ in their proximity to entry sites with more economic activities and better labor force access. Therefore, it is better to compare neighborhoods whose nearby entry sites are similar from the firm’s perspective. To do so, we estimate a propensity score model using candidate entry locations’ demographic and workplace characteristics prior to the entries. We use the model to predict a location’s probability of receiving a high-skilled firm. We then control for a neighborhood’s proximity to nearby potential sites likely to receive entries. We use the remaining variation in a neighborhood’s distance to where the firm actually enters to estimate the causal effects of firm entries. A similar strategy allows us to estimate the effects of firm entries on incumbent residents’ individual outcomes.

Examining the changes in neighborhood characteristics gives us a first order approximation of the welfare impact of the entry on incumbent residents. We find that 6–10 years after the firm entry, house prices and rents in neighborhoods in the treatment group that are within 10 minutes of driving time from the entry increased by 4.3 percent and 1.8 percent, respectively, relative to neighborhoods in the control group that are 20–30 minutes away with similar proximity to nearby potential sites for entry. The average income of workers and residents in neighborhoods in the treatment group increase by 1.5 percent and 2.1 percent, respectively, relative to control neighborhoods. The fact that the increases in rents and wages of workers are comparable in the most treated neighborhoods suggests the incumbent and incoming high-skilled workers who benefit from the firm entry likely

---

<sup>1</sup>We then Bayesian update one’s predicted skill from our baseline model based on the skill mix of one’s Census Block Group of residence in 2010. See Section 2 for more details.

have high willingness to pay for housing services that bid up its price close to the entry. The effects on house prices are larger in cities where land supply is more constrained as a result of geography and land use regulations. In addition, we examine the effects on neighborhood amenities. We do not find economically large and statistically significant effects on the number of establishments per capita for restaurants, nightlife, recreation and private services. Retail amenities like convenience stores and apparel stores decreased by 4 percent in the treatment group, relative to the control. This suggests retail amenities may be vulnerable to rising commercial rents after the entry.

We then study the effects of the firm entry on incumbent residents' financial outcomes by their skill and homeownership status at the time of entry. We track individual outcomes to understand how the welfare impact differs among different demographic groups. Financial outcomes which proxy for well-being give us a more complete picture on the welfare effects, not just through prices changes. Our treatment group consists of incumbent residents living within 10 minutes of driving time from the entry. The control group lives within 20–30 minutes away with similar proximity to nearby potential sites for entry. Our results suggest that high-skilled incumbents, especially homeowners, benefit. We find that high-skilled owners in the treatment group increase non-durable consumption by 1.8 percent 5–8 years later, relative to those in the control group. High-skilled owners are less likely to experience bad credit events such as mortgage delinquencies and debt collections. High-skilled owners benefit from increased wages. They also receive a housing dividend due to the appreciation of their homes, relative to the renters. Hence they can increase consumption and have better financial health. In contrast, for high-skilled renters, we do not find economically large and statistically significant effects on their consumption and financial health.

On the other hand, effects on low skilled renters are inconclusive but suggest they may benefit less or even be harmed. We find neither increase nor decrease in our measures of consumption and financial health for low-skilled renters in the most treated neighborhoods. Since house prices increase close to the entry sites, we expect that low-skilled renters could be harmed. Among low-skilled incumbents, homeowners are better off than renters since the effects of house price increase are neutral to them. Low-skilled owners in the treatment group increase non-durable consumption by 1.8 percent 5–8 years later, relative to those in the control group. We do not find economically large and statistically significant effects on durable-consumption and financial health.

To shed light on the current debate on whether neighborhood gentrification displaces incumbent residents, we also track incumbents' migration decisions. We confirm that high-skilled firm entries can induce migration among incumbent residents that looks like gentrification: low-skilled incumbents are more likely to be displaced further away to lower quality neighborhoods. Low-skilled renters in the treatment group are 5.2 percent more likely to move out of the city. Note that only 17 percent of the low-skilled renters in the control group move out of the city after the entry. This suggests that the number of displaced renters in the treatment group is quite small. In addition, low-skilled renters move to neighborhoods with 2.6 percent lower income per capita and 1.4 percent lower median home value. On average for low-skilled renters, the decrease in neighborhood quality is small. However, for those who get displaced, the decrease in their neighborhood quality is likely

larger, given that income and house prices increase more closer to the entry.

Similarly, low-skilled owners in the treatment group are 6.7 percent more likely to move out of the city. They are also more likely to move to neighborhoods of slightly lower quality. Compared to renters, homeowners are insured against rising rents if they stay in their original neighborhoods. However, we find that low-skilled owners are more likely to move away and increase consumption. This suggests that their wages have not increased as much as the increase in the user cost of owning their homes and the increase in the prices of local non-tradable goods related to housing. Therefore, low-skilled owners find it rational to cash out from their housing assets, move to cheaper neighborhoods, and increase consumption.

To tie together our reduced-form estimates into welfare implications, we estimate a model of worker's choice of home and work locations using a subsample of 129 firm entries announced between 2003–2010 for which we observe commute patterns for incumbent residents. Using the model, we quantify welfare changes due to a representative firm entry on incumbents who live and work within 50 miles of the entry site.

Our results from the model are broadly consistent with the reduced-form evidence. We find homeowners benefit and renters are hurt. High-skilled homeowners benefit the most. After five years, they have an annual welfare increase that is equivalent to a 0.21 percent increase in their initial wages one year prior to the entry. Most of their welfare increase comes from working in places with higher wages and a housing dividend, relative to the renters. Low-skilled renters are hurt the most by an annual welfare-equivalent 0.2 percent decline in their initial wages. This welfare loss per capita is small in comparison to the earning loss due to a typical job displacement which is about 15 percent six years later for workers in Wisconsin ([Couch and Placzek, 2010](#)). Low-skilled renters are hurt because rents increase more relative to increase in low-skilled wages after the entry. Low-skilled workers then live in cheaper neighborhoods further away from the entry. Consequently, they work in places further away from the entry with lower wages.

The welfare effects are especially large for incumbents initially living close to the entry location. Low-skilled renters initially living within 10 minutes of driving time from the entry incur an annual welfare-equivalent loss of 0.62 percent of their initial wages. Further, low-skilled renters who moved away from their original neighborhoods are hurt more than those among stayers, which are consistent with displaced renters being more likely to initially locate in neighborhoods that are hit harder by the entry. The policy implications are that affordable housing units, housing vouchers, or other affordable housing programs could be effective tools to protect vulnerable low-skilled renters and allow them to share the benefits of high-skilled firm entry. Furthermore, property tax could help fund firm subsidies and changes in its design could help mitigate the negative distributional consequences of high-skilled firm entries.

The average welfare change per incumbent seems small because most incumbents are only indirectly affected by the firm that enters. However, the aggregate welfare changes across all incumbents in the metropolitan area for a representative firm entry that promises 1,000 new jobs are still substantial and amount to an annual \$10 million. In comparison, the average one-time

discretionary subsidy for a firm entry is \$107 million for 1,000 promised jobs (Slattery, 2019). The benefits of the entry would exceed the cost of the subsidy in the long run. The welfare effects are heterogeneous across incumbents of different skill and homeownership status. On aggregate, high-skilled owners benefit by an annual \$27.3 million, and high-skilled renters lose by an annual \$4.1 million. Low-skilled owners gain by an annual \$2.6 million, and low-skilled renters lose by an annual \$15.8 million.

Taken together, our results suggest high-skilled firm entry in the U.S. has benefited high-skilled incumbent residents, particularly those who are homeowners. In contrast, low-skilled incumbents on average benefit less. In particular, low-skilled renters are harmed. However, the typical high-skilled firm entry in the U.S. has small welfare consequences on the average incumbent resident. Our findings reaffirms the classic intuition among economists from a large literature of spatial equilibrium models starting from Rosen (1979) and Roback (1982) that individuals make behavioral adjustments which could offset the impact of local price changes. A caveat is that the welfare consequences of some extreme cases of large high-skilled firm entry could be much larger due to more extreme changes in local prices. Moreover, there could be multiple large firm entries into an area within a short period of time that amplify the welfare consequences, whereas our welfare calculation is for a single representative entry in our sample. Both heterogeneity of welfare effects across entries and measuring the multiplying effects of entries are potential future points of research.

## 1.1 Related Literature

This paper builds on an extensive literature on the effects and incidence of spatial treatment.<sup>2</sup> The specific part of the literature that we are most closely related studies the effects of total factor productivity growth related to firm entry.<sup>3</sup> Examples include Greenstone et al. (2010) and Monte et al. (2018) on agglomeration spillovers in counties that attract a large plant opening and Currie et al. (2015) on local dis-amenities of toxic plant operation. Recently, Hornbeck and Moretti (2018) study the effects of city-wide productivity gains in manufacturing. Eckert et al. (2020) study how high-skilled service industries are responsible for the urban bias in economic growth in the U.S. Our paper is different on a number of important dimensions. Previous papers on firm entries focus on studying the effects of entry on county or neighborhood level outcomes. We are interested in the differential welfare effects of a firm entry on incumbent residents. This requires us to observe incumbents' movements along with their individual outcomes. We bring to bear novel microdata of individual migration histories linked to housing characteristics and financial outcomes. We provide the first set of empirical evidence on how incumbents are impacted differently depending on their skill and homeownership status. Second, our identification strategy improves on the commonly used ring analysis with controls for neighborhood characteristics in studying spatial treatment effects.

---

<sup>2</sup>Examples include Busso et al. (2013) on neighborhood economic development programs, Autor et al. (2013) and Chetverikov et al. (2016) on trade shocks, Bartik (2018) on economic development incentives, Diamond and McQuade (2019) on effects of low-income housing tax credits on local neighborhoods, and Yagan (2019) on the Great Recession.

<sup>3</sup>A sub-part of the literature related to our work looks at the local impacts of hydraulic fracturing activities. Examples include Muehlenbachs et al. (2015), Feyrer et al. (2017) and Bartik et al. (2019).

We address the endogeneity of firm’s entry location by controlling for a neighborhood’s proximity to potential sites for entry predicted by a propensity score model. This allows us to match a rich set of neighborhood characteristics between neighborhoods close to and far away that are correlated with the likelihood of attracting a large high-skilled firm in a parsimoniously way.

Our analysis also sheds light on the consequences of neighborhood gentrification and its mechanisms. Our paper relates to a broad literature across disciplines studying the effects of gentrification. Most previous studies have focused on whether gentrification leads to displacement of socio-economically disadvantaged local residents as the main outcome of interests, and find little evidence of displacement (Vigdor, 2002; Freeman and Braconi, 2004; Freeman, 2005; McKinnish et al., 2010; Ellen and O’Regan, 2011; Ding et al., 2016; Dragan et al., 2020; Martin and Beck, 2018). On the other hand, Martin and Beck (2018) find that gentrification directly displaces renters, but not homeowners. However, displacement by itself does not necessarily imply any welfare loss. A recent exception is Brummet and Reed (2019) who use longitudinal Census microdata to study the effects of neighborhood gentrification on the well-being of original residents. They find gentrification increase out-migration, but the residents who stay benefit from improvements in neighborhood quality.

We contribute to this literature in few ways. First, our analysis goes beyond studying displacement, and estimates causal effects on incumbent residents’ financial outcomes and neighborhood quality, which approximate their well-being. Second, previous papers mostly use changes in neighborhood observables such as the change in share of residents with a college degree to detect whether a neighborhood is undergoing gentrification. However, neighborhood level changes could be triggered by different causes across space and time, which makes it difficult to generalize the effects of gentrification across places. The endogeneity of neighborhood changes poses additional challenge for estimating the causal effect of gentrification on local residents. We instead focus on high-skilled firm entries as a potential trigger, and find they could induce migration response among incumbent residents that look like gentrification: low-skilled incumbents are more likely to be displaced out the city and move to neighborhoods with lower quality. Third, our identification strategy which controls for a host of neighborhood characteristics that influence firm’s entry location, combined with the use of a high-skilled firm entry as a large concentrated shock that dominates many underlying trends in the entry area, allows us to more credibly estimate causal effects of the entry.

Our model contributes to a broad literature using structural models to evaluate the welfare effects of neighborhood demographic changes. The literature has examined different triggers of neighborhood demographic changes, e.g., a city’s division and reunification (Ahlfeldt et al., 2015), new urban transit infrastructure (Severen, 2018; Tsivanidis, 2019), improvements to neighborhood labor market opportunities (Baum-Snow et al., 2019), the rise of home-sharing (Almagro and Dominguez-Iino, 2020; Calder-Wang, 2019) and ride-sharing platforms (Gorback, 2020), and various causes of urban gentrification such as rising incomes at the top of the income distribution (Couture et al., 2019), changing preferences of young college graduates for amenities (Couture and Handbury, 2019), and changing preferences for value of time (Su, 2019). We add to this literature



by studying how high-skilled firm entries lead to neighborhood demographic changes and their welfare consequences. Specifically, we build a structural model, which features choices of residential and workplace locations within a city and heterogeneity in workers’ preferences. We use the model to evaluate the welfare incidence of high-skilled firm entries on incumbent residents by their skill group and homeownership status.

This paper also relates to the literature on state and local firm entry subsidies, including [Giroud and Rauh \(2019\)](#), [Bartik \(2017\)](#), [Suárez Serrato and Zidar \(2016\)](#), and [Fajgelbaum et al. \(2019\)](#). [Slattery \(2019\)](#) examines the bidding process for state firm subsidies and tax incentives and [Slattery and Zidar \(2020\)](#) provides an overview of many types of incentive policies. Our contribution is to quantify the benefits of firm entries to the city and the distributional consequences to be considered when local governments attract firms with financial incentives.

The remainder of the paper is structured as follows. Section 2 describes our data set, and Section 3 describes our empirical approach. Then Section 4 describes our reduced-form results. Section 5 introduces a model of worker’s home and work location choices to evaluate welfare incidence. Section 6 concludes.

## 2 Data

For the analysis in this paper, we merge data from multiple sources to compile a rich panel data set on 45 million individuals who are potentially affected by at least one of the 391 firm entries we study. We summarize the key details here, and Appendix C provides many more details on each part of the data.

### 2.1 Data and Sample Construction

#### 2.1.1 Sample of Firm Entries

The firm entries in our sample come from Conway Analytics. Conway is a company that offers services to governments and companies that helps them decide where to locate new establishments. Conway produces the *Site Selection Magazine*, used in [Greenstone et al. \(2010\)](#)’s paper on large plant openings. Conway Analytics provides data on large corporate expansion projects, and we obtained 3700 records from 1990–2010 where a firm announced it will open an establishment in the U.S. that will create at least 500 new jobs ([Conway \(2018\)](#)). For each establishment entry, Conway provides the investment amount and the NAICS industry code at the establishment level. From these 3700 firm entries, we use the NAICS codes to filter to 391 entries where the establishment was in the high-technology industry, using the definition provided by the [National Science Foundation \(2020\)](#). Furthermore, we manually verify that each firm entry employs mainly high-skilled workers by looking at their online job postings; for example, for Amazon, we are interested only in establishments where software engineers work and exclude establishments that are warehouses or distribution centers. We keep cases where the firm announced 500 new jobs but substantially fewer occurred in practice after entry. We drop cases where the firm announced an entry, but the



entry never happened in reality. We also drop firm entries that have the same closest CBSA and are within 5 years of each other. We link each firm to its address using Google Maps and then geocode each entry and map to its Census Tract.

Figure 1 plots each firm entry scaled by the promised number of jobs, so larger entries are represented by larger circles. Most of the entries occur in metropolitan areas. The entries are spread out across many states and regions of the US.

Table 1 provides summary statistics for the geocoded sample of 391 firm entries from Conway. All entries promised to employ at least 500 workers, and most of the jobs are high-skilled. There is a substantial amount of variation in the sample with respect to location and neighborhood characteristics: distance to the closest downtown city centers (as proxied by the CBSA center), median household income, renter share, and college share. Jobs promised averages 1,091 jobs, which indicates these establishments are quite large relative to the population of all establishments. Investment amount averages \$245 million, with a wide range from \$8 million in the 10th percentile to \$700 million in the 90th percentile.

### 2.1.2 Sample of Individuals

From a marketing dataset, Infutor<sup>4</sup>, we obtain address histories for a sample of 45 million incumbent residents between 20 and 65 years old living within 30 minutes of driving time from each of the 391 firm entry locations as of December 31 of the year prior to the firm entry announcement. Infutor provides the exact street address, the month and year in which the individual lived at an address, the name of the individual, and some demographic information including age and gender. We geocode the street addresses to map it to 2010 Census Blocks in order to determine which neighborhoods incumbent residents lived in along their migration histories before and after the firm entry.

We merge Infutor to property records provided by CoreLogic, which provides tax assessor and transactions data for the near universe of residential properties in the U.S. These data provide us with property address and a variety of characteristics, such as the use-code (single-family, multi-family, etc.), the year the building was built, and the number of units in the structure. For each property, the data also provides us with the transaction history since 1990, which includes sales prices, buyer and seller names, and the dates of sales. By comparing names from Infutor to the owners of the property in CoreLogic, we determine whether an incumbent resident is a homeowner or renter at each point in time, as detailed in Section 2.2.1. We further predict individual’s education level using a rich set of demographic and housing characteristics from our Infutor-CoreLogic linked data, as detailed in Section 2.2.2.

For a representative subsample of 8 million individuals, we obtain their credit reports from

---

<sup>4</sup>Infutor is a company that aggregates address data from many sources including phone books, voter files, property deeds, magazine subscriptions, credit header files, and other sources. Bernstein et al. (2019) show Infutor covers about 80% of the U.S. adults in 2000, and captures 99% of the cross-sectional variation in county-level population in the Census.

credit bureau TransUnion (TU) from 2000 to 2018.<sup>5</sup> TU used SSN identifiers and names to link individuals to financial outcomes and returned an anonymized dataset retaining the demographic and geographic information. Variables from TU that reflect an individual’s financial health include credit score, credit events such as mortgage delinquencies, bankruptcies, and debt collections, and various loan balances.

Thus, we have time-varying individual information for renter/owner status and financial outcomes, and we have time-invariant information for birth year, gender, and education. Panel A in Table 2 summarizes the characteristics of our sample of individuals. We see the average incumbent resident in our sample is about 43 years old in the year of firm entry announcement and has lived at his current address for 5.5 years. We also see high-skilled incumbents and homeowners made up 36 percent and 53 percent of the sample prior to the firm entry. Four years prior to the firm entry announcement, 8.8 percent of the incumbents lived more than 20 miles from their addresses at the time of announcement. The incumbent residents on average have a credit score of 682<sup>6</sup>, and 3.3 bank revolving accounts including credit and debit cards.

### 2.1.3 Construction of Neighborhood-Level Outcomes

Our sample of neighborhoods are made of Census Tracts of ZIP Codes within 30 minutes of driving time from each of the firm entry location. We obtain outcomes for Census Tracts from 1990, 2000, 2010 Decennial Censuses and 2005–2018 5-year American Community Survey (ACS) aggregate statistics from [IPUMS NHGIS \(2020\)](#). Census Tracts are harmonized to 2010 Census Tract boundaries following the method released by LTDB ([Spatial Structures in the Social Sciences \(2020\)](#)). For house prices at the Census Tract level, we build a hedonic house price index using housing transactions from CoreLogic, as detailed in Appendix C.3. We obtain outcomes for ZIP Code Tabulation Area (ZCTA) codes in 2010 boundaries, which standardized ZIP Codes, from a few data sources. We obtain annual ZIP Code income from the IRS Statistics of Income Tax Stats. ZIP Code Business Patterns (ZBP) provides counts of business establishments by employment size buckets and total payroll across time for each ZIP code. Panel B in Table 2 summarizes the characteristics of our sample of neighborhoods. The average annual income of residents and workers in those neighborhoods are 73K and 42K in 2010 dollars. The house prices are on an upward trajectory and on average have increased by 6 percent in the four years prior to the firm entry.

### 2.1.4 Commute Patterns

In order to estimate a model for worker’s choice of home and work locations which we use to evaluate welfare incidence, we need to observe the share of incumbent residents by skill group

---

<sup>5</sup>Table A4 shows whether Infutor panelists are matched to TU credit reports is not correlated with where they are initially located when the firm entry is announced.

<sup>6</sup>Credit scores are from VantageScore 3.0 model, which was developed by the 3 major credit bureaus including Experian, Equifax, and TransUnion. VantageScore 3.0 uses a range between 300 and 850. A VantageScore above 660 is considered good, while a score above 780 is considered excellent.

who choose each combination of residential and workplace neighborhoods. The LEHD Origin-Destination Employment Statistics (LODES) provides Tract-to-Tract worker commuting flows from 2002–2017. For each choice of residential and workplace Tract combination, LODES provides the number of workers who make that choice. We then predict what share of the commuters are high-skilled by building a machine learning model which takes as input the characteristics of each pair of residential and workplace neighborhoods, as well as characteristics of workers commuting between them using a few data sources including the Census/ACS, ZBP, and LODES. Finally, we calculate the choice probabilities of incumbent residents by skill group for different home-work locations surrounding the firm entry site one year before and five years after the entry. Details of data implementation for the model estimation including validations of our commute pattern model, and how we measure neighborhood observables including wages and rents are provided in Appendix C.5.

## 2.2 Construction of Important Variables

### 2.2.1 Homeownership Status

The effects of a high-skilled firm entry will be different for incumbent homeowners and renters. Unlike renters, homeowners are perfectly insured against the risk of rising rents after the firm entry. We therefore want to distinguish owners from renters and impute homeownership status for incumbent residents in our individual sample. Appendix C.1 provides the full details.

To impute home ownership, we first construct an address-year panel with property owner names of each address in each year using both tax records and transactions data from CoreLogic. Next, we compare the last names of individuals from Infutor and the last names of property owners from CoreLogic in each year to determine whether each incumbent resident is an owner or renter in each year since 1990. Figure B2 validates the homeownership imputation by comparing our homeownership rate at the Census Tract level with the homeownership rate from the 2000 Census, 2008–2012 5-year ACS, and 2013–2017 5-year ACS. Our homeownership can explain more than 95% of cross-sectional variation among different years. We can be reasonably confident that our measure of homeownership is of high-quality.

### 2.2.2 Education/Skill

The effects of a high-skilled firm entry on an incumbent resident also depends on his skill group. High-skilled wages likely increased more than low-skilled wages in nearby areas because high-skilled workers are more directly employable by the entering firm. We classify individuals as high-skilled or low-skilled based on their education level: those with at least a four-year college degree are high-skilled and those without are low-skilled. While education and skill are not identical, education may be viewed as a proxy for skill, and we will use the two terms interchangeably in this paper. Full details of skill imputation are provided in Appendix C.2

To predict education, we first build a model using a flexible logit model that predicts whether

a person has a college degree or above based on a rich set of individual demographic and housing characteristics that we observe both in 2010 ACS and our Infutor-CoreLogic linked data. These individual characteristics are highly correlated with education level. For example, an older worker living in a smaller house that is rented is more likely to be low-skilled. Next, we use the model to predict out-of-sample on our sample of incumbent residents based on their observed demographic characteristics and housing characteristics of their address in Dec. of 2010. Because individuals prefer living close to people with similar education levels, if an individual is observed to reside in a predominantly high-skilled neighborhood, then we should be more confident that he is a high-skilled worker, and vice versa. Hence, we Bayesian update each individual’s baseline high-skilled probability with the share of high-skilled working-age population by gender from their 2010 Census Block Group.

We validate the skill imputation by checking the skill distribution of our sample against the Census. Table A3 Column 1 shows the skill distribution of our sample of incumbents in 2010, which agrees well with the 2010 ACS measure of skill distribution, reported in Column 2. Column 4 shows that individuals who we classify as high-skilled in 2010 live in neighborhoods with a higher share of high-skilled workers in 2015, which is consistent with what one would expect from some degree of continued skill sorting. This is an out-of-sample check since we use their 2010 address in our skill imputation. Thus, we can be reasonably confident that our measure of skill is of good quality.

### 2.2.3 Driving Time

We quantify the treatment intensity of each neighborhood by the firm entry using 1990 level historical driving time. Neighborhoods with shorter driving distance to the firm’s entry location have better access to jobs near where the firm enters. Hence high-skilled workers newly employed by the entering firm would prefer living in those neighborhoods. Consequently, incumbent residents living there are more affected as they likely experience larger changes in local house prices and amenities. We use historical travel time rather than contemporaneous travel time to measure neighborhood’s treatment intensity because today’s travel time could have responded to the firm entry.

To calculate historical travel time, we estimate a speed model using morning and evening commute trips from the 1995 National Household Travel Survey (NHTS), similar to Couture (2016) and Su (2019). We apply sampling weights for trips to make sure trips in our sample are nationally representative. We only use an 80% sample of NHTS trips to estimate the model, with the remaining 20% of the data held out as a test sample to evaluate model fit. Then we use the model parameters to predict 1990 travel speed and travel time, using 1990 Census characteristics of neighborhoods. See Appendix C.4 for more details.

In Figure B3, we evaluate the accuracy of our speed model in two ways using the 20% test sample, and the 2019 contemporaneous travel time from Google Maps. These two methods validate that our speed model yields accurate predictions.

### 3 Empirical Approach

After a high-skilled firm entry, the changes in nearby neighborhoods and outcomes of incumbent residents living in them could be attributed to three main channels: 1) The direct effect of the jobs created by the large high-skilled firm entry, 2) the indirect general equilibrium (GE) effects due to ancillary jobs created by the entering firm as well as jobs created by other complementary firms that follow the large high-skilled firm, and 3) the entry area-wide general equilibrium effects as a result of both the large high-skilled firm entry and pre-existing firms prior to the entry. In the third channel, some of the jobs would have been created by the pre-existing firms had no high-skilled firm entered, since the whole area could be on an upward growth trajectory.

We propose a strategy to identify the causal effects of the firm entry on neighborhood outcomes and outcomes of incumbents living in affected neighborhoods that are due to neighborhoods' differential amount of exposure to the firm entry shock. This strategy allows us to capture the combined effects of the entry that are due to both (1) and (2), while differencing out the area-wide GE effects due to (3).

To estimate the effects of firm entries due to neighborhoods' differential exposure, a simple strategy would be to compare neighborhoods close to and far away from the firm entry location. However, a high-skilled firm picks an entry location based on its desirability to the firm. Thus, neighborhoods close to the desirable locations for the firms may be unobservably different from those far away. To overcome this source of bias, our solution is to control for these unobservables by controlling a neighborhood's proximity to desirable locations within the same metropolitan area as the actual entry location. We then use the remaining variation in a neighborhood's distance to where the firm enters to identify the causal effect we defined.

#### 3.1 Predict Desirable Neighborhoods for High-skilled Firm Entry

In order to execute our identification strategy, we first need to know where are the desirable locations for high-skilled firm entry in general. When a high-skilled firm chooses a location to enter, its desirability is likely influenced by many demographic and workplace characteristics of neighborhoods around the entry location. For example, a high-skilled firm may be more attracted to enter a place with a young workforce. Benefits of agglomeration may lead a high-skilled firm to locate close to other pre-existing high-skilled industries that are complementary to its production. We use a propensity score model to parsimoniously summarize a location's desirability for attracting high-skilled firms. Here we summarize the key details, and Appendix D provides more details.

For each of the 410 firm entry announcements in our sample,<sup>7</sup> we draw a ring with a 50-mile radius around the announced entry location. The 50-mile ring roughly corresponds to the distance it takes to travel out of the metropolitan area. For each ZIP Code inside the 50-mile ring,<sup>8</sup> we predict its probability of receiving a high-skilled firm entry by estimating a propensity score model

---

<sup>7</sup>The sample consists of 391 actual entries and 19 announced entries that never occurred ex post. The main analysis uses only the 391 entries that did occur.

<sup>8</sup>ZIP Codes are defined by 2010 ZIP Code Tabulation Areas (ZCTAs), and are time-invariant in their boundaries.

using a Lasso-Logistic regression. The advantage of Lasso regularization is that it avoids overfit by penalizing the number of variables. The regression sample consists of all ZIP Codes within 50 miles of each entry location, and a ZIP Code is assigned an indicator equal to 1 if it was ever home to one of the 410 entries and 0 otherwise. ZIP Codes associated with 90% of the 410 firms comprise the training data, and ZIP Codes associated with 10% of the 410 firms are held out as the testing data.

The propensity score model takes as input a rich set of 1990 level demographic and workplace characteristics of not only each ZIP Code itself, about also characteristics of its surrounding neighborhoods. Specifically, for each ZIP Code, we construct successive “ring” measures of 1990 level average features of its surrounding ZIP Codes within 0–5 miles, 5–10 miles, 10–20 miles, and 20–30 miles, and use them as regressors. We use 1990 level neighborhood characteristics in order to predict the desirability of a location for high-skilled firm entry, prior to any firm entries in our sample. We include in our model ZIP Code level demographic characteristics such as age and education composition of residents, indicators for economic conditions including median household income and unemployment rate, as well as housing characteristics including median home value and median rent. We also include ZIP Code level workplace characteristics such as the average wage of workers, pre-existing job densities by industries, particularly high-tech job densities, employment-to-population ratio which measure how commercial vs. residential a neighborhood is, and distance to the closest Central Business District (CBD) which measures a neighborhood’s urban/rural status. Demographic characteristics might have different influence on a location’s desirability depending on its business concentration and its urban/rural status. Hence, we also allow employment-to-population ratio and distance to the closest CBD to interact with the set of demographic characteristics. Table A16 and Table A17 contain the complete list of demographic and workplace characteristics in the model. Table A5 shows the estimated parameters of the propensity score model. Noticeably, the probability of receiving a high-skilled firm entry in a ZIP Code is positively correlated with nearby job densities in high-tech industries, employment-to-population ratio, and variables that indicate the local workforce is young.

In Figure B5, we show that the estimated propensity score model has good out of sample fit by plotting each ZIP Code’s actual probability of being home to an entry against the predicted probability of having an entry from our model. The linear best fit for the test data has a slope of 1.113, which is similar to the slope of 1.105 for the training data with a significance level of 99%. This indicates that the model has similar predictive power in both the training and testing data.

### 3.2 Main Specification for Neighborhood and Individual Outcomes

Figure 2 shows an example that illustrates our estimated propensity score model. It plots the predicted probability of receiving a high-skilled firm entry for ZIP Codes within 50 miles of the entry location of Pfizer in Ann Arbor, Michigan. Pfizer makes pharmaceutical products, and it announced the entry of its R&D facility in 2001, which promised to hire 600 new employees with over \$700 millions of investment. The red areas, which have the highest propensity scores, are

correlated with where cities are located, and the green areas have the lowest propensity scores. The cities of Lansing and Troy both have neighborhoods with relatively high propensity scores among neighborhoods in the metropolitan area, but lower propensity scores than some sites in Ann Arbor. From Pfizer’s perspective, potential sites in Lansing and Troy could have been desirable as entry locations as well.

Using Pfizer’s entry to illustrate our identification strategy, we compare neighborhoods close to and far away from the actual entry location in Ann Arbor, while controlling for neighborhood’s proximity to high-propensity areas in Lansing and Troy. To identify the effects of the firm entry due to differential exposure, we use the remaining variation in a neighborhood’s distance to where the firm actually enters. A caveat of this strategy is that we can only capture the relative differences between neighborhoods close to and far away due to the direct effects of the entry and indirect GE effects which are correlated with the distance to the firm entry. This is not the total effects of the entry which would also include the area-wide GE effects.<sup>9</sup>

One concern with this strategy is that whether there is any variation left over after we control for proximity to desirable locations for entry. If high-skill firm entries only depends on the desirability of a location as predicted from our model, there would be no “good” source of variation to isolate in order to identify the treatment effects. In reality, firms choose between many pretty similar locations with high propensity scores, and there is a lot of idiosyncrasies in why they end up in a particular neighborhood. In Pfizer’s case, above and beyond the neighborhood characteristics that we control for in our propensity score model, Pfizer may prefer Ann Arbor over Lansing because the labor skills it demands are better matched by those supplied by the medical school at the University of Michigan at Ann Arbor than by the medical school at Michigan State University in Lansing. Such idiosyncrasies contribute to the quasi-random variation we leverage for identification. Furthermore, across our 391 firm entries, they often chose locations with high propensity scores, but some entries landed in locations that do not have high propensity scores, relative to other places in the same metropolitan area. This contributes to another useful source of variation.

To execute our identification strategy, for each firm entry, we define an entry zone  $k \in \{1, \dots, K\}$  made of neighborhoods within 30 minutes of 1990 level driving time from the actual entry location.<sup>10</sup> We divide neighborhoods within each entry zone into three rings. The inner ring of neighborhoods 0–10 minutes away from the firm entry is the most treated group, the 10–20 minute ring is the second most treated group, and the outer ring 20–30 minutes away from the entry is the control group, which serves as a comparison group in order to difference out area-wide GE effects.<sup>11</sup>

---

<sup>9</sup>An alternative strategy is to use comparable entry locations in runner-up cities as a control group in the spirit of [Greenstone et al. \(2010\)](#). Though we do not have data on runner-up cities for most of the entries in our sample, it may be feasible to construct runner-up cities via a propensity score model or synthetic controls. However, the drawback is that finding suitable control cities is inherently difficult for large metropolitan areas that are outliers by definition, and even more difficult to do so in a scalable way for our sample of 391 firm entries. Even if runner-up cities could be found, we still have to predict the exact entry locations using a propensity score model. However, unobserved differences in state and city governments’ investment incentives are difficult to fully control for when we compare desirable entry locations across cities.

<sup>10</sup>This roughly corresponds to a ring of 25 miles in radius.

<sup>11</sup>Figure [B4](#) shows the distribution of driving time from the firm entry location to the centroid of each Census



To evaluate the effects of the firm entry on neighborhood level outcomes, we use an event study design with the following exact specification:

$$Y_{i,k,t} = \alpha_0 + \sum_g \sum_{\tau} \delta_{\tau}^g D_{\tau,k} \times g_{i,k} + \eta_{i,k} + \phi_{h^1,h^2,k,t} + \varepsilon_{i,k,t}, \quad (1)$$

where  $Y_{i,k,t}$  is the outcome of neighborhood  $i$  in the entry zone  $k$  in year  $t$ .  $D_{\tau,k}$  is an indicator for  $\tau$  years relative to the announcement year, and  $g_{i,k}$  is an indicator for whether neighborhood  $i$  in entry zone  $k$  belongs to one of the two treatment groups, indexed by  $g \in \{1, 2\}$ .  $\delta_{\tau}^g$  measures the causal effect  $\tau$  years after the firm entry for treatment group  $g$ , relative to the control group. We normalize the coefficient of the year prior to announcement,  $\delta_{-1}^g$ , to zero.  $\eta_{i,k}$  are Census Tract by entry zone fixed effects,<sup>12</sup> which control for time-invariant differences in neighborhoods that are correlated with the likelihood of attracting a high-skilled firm nearby.

We further control for a neighborhood's proximity to potential sites for high-skilled firm entry as predicted by our propensity score model. To do that, we divide our treated and control neighborhoods into groups such that within each group, treated and control neighborhoods are comparable in their closest distances to potential sites likely to receive high-skilled firm entries. Hence the neighborhoods we compare are similar in all kinds of demographic and workplace characteristics that are correlated with the likelihood of attracting a high-skilled firm entry. Specifically, let  $\mathbf{Z}_{i,k} = [Z_{i,k}^1, Z_{i,k}^2]$  denote the vector of the closest distances from neighborhood  $i$  in entry zone  $k$  to ZIP Codes within 50 miles from the entry location in the quintiles,  $Q_1$  and  $Q_2$ , that have the highest propensity scores. We then split neighborhoods into bins based on their closest distances to potential entry sites in  $Q_1$  and  $Q_2$ . In the main specification, we split  $Z_{i,k}^1$  and  $Z_{i,k}^2$  into bins defined by 0–5, 5–10, 10–20, and 20+ miles, and denote the resultant bins  $h_{i,k}^1$  and  $h_{i,k}^2$ . We then allow neighborhoods in the same group as defined by bins  $h_{i,k}^1$  and  $h_{i,k}^2$  to have their own time trend by adding a set of distance bins by entry zone by year fixed effects, denoted by  $\phi_{h^1,h^2,k,t}$ . Note these fixed effects subsumes the entry zone by year fixed effects that we otherwise need to control for local time trends. Standard errors are clustered at the entry zone level to account for potential correlation of outcomes of neighborhoods treated by the same firm entry. For robustness check, we also include controls for a neighborhood's proximity to ZIP Codes with high prior density of high-tech employment. Hence, the identification is more explicitly based off comparing neighborhoods with similar access to pre-existing high-tech jobs in nearby places.

To evaluate the effects of the firm entry on incumbent residents' outcomes, we define treatment groups as incumbents who were living in the inner and middle rings when the firm entry was announced, the control group consists of incumbents living in the outer ring. we estimate a similar

---

Tract within 40 minutes of driving time. It shows there are a sufficient number of Tracts within 30 minutes of the firm entry. The distribution of Tracts is smooth across the range of driving time considered, so there does not exist discontinuities around 30 minutes. There are also a sufficiently large number of neighborhoods in each ring so that we have enough power to measure outcomes accurately.

<sup>12</sup>If neighborhood outcome is measured at the ZIP Code level, then we include ZIP Code by entry zone fixed effects.

event study design with the following exact specification:

$$Y_{d,i,k,t} = \alpha_0 + \sum_g \sum_{\tau} \delta_{\tau}^g D_{\tau,k} \times g_{d,i,k} + \eta_{i,k} + \phi_{h^1,h^2,k,t} + \psi_{a,n,k,t} + \varepsilon_{i,k,t}. \quad (2)$$

Each observation is defined by a combination of Census Tract, skill level, homeownership status, entry zone, and calendar year. Thus,  $Y_{d,i,k,t}$  is the average outcome of incumbents in calendar year  $t$  who belong to a skill and homeownership group  $d$ , and were living in Census Tract  $i$  in entry zone  $k$  when the firm entry is announced. In addition to the same set of fixed effects in the specification for neighborhood level outcomes, we also control for two demographic status, age  $a$  and tenure  $n$ . This is because younger people tend to have higher mobility and people with longer tenure at their address when the firm entry is announced tend to move less due to having built up more neighborhood capital. We then allow each group of incumbents defined by age and tenure to have their own time trend. Specifically, we add a set of fixed effects,  $\psi_{a,n,k,t}$ , which denotes the interaction of dummies for age group  $a$  and tenure group  $n$  with dummies for each entry zone  $k$  and calendar year  $t$ .<sup>13</sup> Individual outcomes are demeaned to account for individual fixed effects. Regressions are weighted by the number of incumbents in each neighborhood.<sup>14</sup> Standard errors are clustered at the entry zone level.

The identifying assumptions are conditional on the fixed effects, (i) the treatment and control groups have parallel trends prior to the firm entry, and (ii) there are no omitted shocks occurring before the firm entry that are correlated with the treatment group assignment. As we mentioned before, our identification strategy differences out the city-wide GE effects. A concern is that these GE effects are different across our treatment and control groups prior to the entry. For example, if after including our controls for a neighborhood's proximity to potential sites for entry, there are still differences in job growth between the inner and outer rings due to the opening of other large firms around the time firms in our sample are opening, then our estimated effects could be biased.<sup>15</sup> The fact there are no pre-trends is consistent with negligible differences between treatment and control groups in terms of outcomes prior to the firm entry due to these GE effects. Note the lack of pre-trends is consistent with no omitted shocks, it does not test for a violation of this assumption.

---

<sup>13</sup>We define two age groups: young vs. mature, where young incumbents are those younger than 40 when the firm entry is announced. We define two tenure groups: short vs. long, where incumbents with a short tenure have stayed at their addresses for less than 4 years when the firm entry is announced.

<sup>14</sup>Due to the computational burden imposed by our large sample size, we choose this specification which collapses outcomes for incumbents in each neighborhood in our treatment and control groups, instead of running an individual-level event study.

<sup>15</sup>We manually examine a random subset of 10 entries and find that for only 2 of the 10 cases, it looks like another large firm entered just before, or there was generally a lot of economic development existed prior to firm entry. See Appendix C.6 for details.

## 4 Reduced Form Results

### 4.1 Validation of Propensity Score Model

We validate our propensity score model using a simple event study design. The treatment group are ZIP Codes that actually received an entry in our sample, the control group are ZIP Codes that did not receive any entry in our sample, but among the top quintile of ZIP Codes within 50 miles of each actual entry location that have the highest predicted probability of receiving a high-skilled firm entry. Appendix D provides the full specification and more details including additional robustness checks. Our purpose is two-fold: (1) Relative to the other top potential sites, the treated sites should have no differential growth of high-skilled jobs prior to the actual entry. (2) A source of bias could arise because we do not observe the universe of firm entries in our sample. When we compare outcomes of residential or workplace neighborhood close to and far away from the treated sites, controlling for their proximity to potential sites for entry, if the neighborhoods far away are also receiving a comparable amount of high-skilled jobs growth nearby as neighborhoods close to the treated sites, our estimated effects could be downward biased. Hence we want to confirm that the treated sites experience significantly higher growth of high-skilled jobs after the entry, relative to the potential sites.

Figure 3 plots the event study coefficients. The lack of pre-trends in the event study coefficients confirms there is no differential growth of high-skilled jobs prior to the firm entries between locations that received the entries and other top potential locations within the same metropolitan area. After the entry, Figure 3 and Table A6, Panel A both show that the ZIP Codes which received an entry in our sample of firm entries on average received a significantly higher amount of high-skilled job growth, relative to the other potential sites in the same metropolitan area. On average, they received an additional increase of 0.75 high-tech establishments of size 500+, and an additional increase of 2.48 high-tech establishments of size 100–499, relative to the other potential sites 6–10 years after the firm entry announcement. These evidence suggest the vast majority of the firm entries in our sample did reach their target capacity of 500+. They also brought in other smaller and complementary high-tech establishments following our large firm entries as well. We perform robustness checks in Table A6, Panel B by only comparing ZIP Codes which received an entry in our sample with other top potential sites within the same county as the entry location. This further controls for potential unobserved heterogeneity between different counties in the same metropolitan area, such as differences in corporate tax incentives. In Table A6, Panel C, we further control for dynamic aspects of neighborhood changes that could affect firms’ location decisions. We add changes in neighborhood demographic and workplace characteristics in the decade prior to each firm entry for neighborhoods surrounding each potential entry location to our propensity score model. The estimated event study coefficients are highly robust in both alternative specifications for model validation.

## 4.2 Effects on Neighborhood level Outcomes

We now switch perspective from treating neighborhoods as potential sites for firm entry to examining the effects of the firm entry on nearby residential or workplace neighborhoods. Our treatment groups are neighborhoods within 10 minutes of driving time from the entry in the inner ring and those 10–20 minutes away in the middle ring. The control group are neighborhoods 20–30 minutes away in the outer ring.

Figure 4 plots the event study coefficients of estimating equation (1) for four neighborhood level outcomes: house price indices and median rent at Census Tract level,<sup>16</sup> average income of workers in a workplace ZIP Code, and average income of residents in a residential ZIP Code. There is little evidence of differential trends in these neighborhood outcomes between the treatment group and the control group prior to the firm entry announcement. After the entry, there is clear evidence that the firm entry leads to higher house prices and rents, and higher income in the inner ring, relative to the outer ring. There are no economically large or statistically significant effects on the middle ring. Hence, most of the treatment effects due to differential exposure to the firm entry shock are concentrated in the inner ring. We thus focus our subsequent discussion on the effects on the inner ring.

Figure 5 summarizes the average treatment effects for the inner ring, relative to the outer ring, 6–10 years after the entry. All effects are presented in percent terms, so they are scaled by the mean of the control group. We find that house prices and rents in neighborhoods in the inner ring increased by 4.3 percent and 1.8 percent, respectively, relative to neighborhoods in the outer ring. Average income of workers in a workplace ZIP Code in the inner ring increase by 1.5 percent, relative to the control group. We also find the average income of residents in a residential ZIP Code in the inner ring increase by 2.1 percent, relative to the control. This is likely due to the residents who choose to live in the inner ring after the entry are those whose wage increases are high enough to afford the rising house prices and rents there. Focusing on the comparison between rent and workplace income which are both flow variables, our results indicate the increases in rents and wages of workers are comparable in the most treated neighborhoods. A possible explanation is that the incumbent and incoming high-skilled workers who benefit from the firm entry likely have very high willingness to pay for housing services, and they bid up the prices of housing services close to the entry.

In Table 3, we examine whether the treatment effects on neighborhood level house prices, rents, and income are heterogeneous across cities with different initial conditions and across different industries. We estimate equation (1) for each firm entry individually. Figure B10 plots the treatment effects for the inner ring relative to the outer ring, 6–10 years after the entry for each individual case. We then regress the case-by-case treatment effects on four variables: (1) land-supply elasticity in the CBSA of the entry location. This measure is taken from [Saiz \(2010\)](#) and takes into account of both land availability due to geographic constraints and strictness of land-use regulations. (2) the share of high-skilled workers who live in each entry zone prior to the entry. (3) commuting

---

<sup>16</sup>Appendix C.3 documents how we construct the Tract level house price indices using CoreLogic transaction data.

openness defined as the share of residents living in each entry zone who work outside their entry zone prior to the entry.<sup>17</sup> (4) 6 industry category dummies of the firm entry which are pharmaceutical, energy and equipment, hardware technology, software technology, finance and insurance, and other services.<sup>18</sup> We find the effect on house prices is smaller when the firm enters a city where land supply is less constrained. A standard deviation increase in land supply elasticity decreases the effect on house prices in the inner ring in the long run by 2.56% ( $1.11 \times -0.0231$ ), relative to the outer ring. This difference is significant at the 95% level. We also find suggestive evidence that technology firms that make software and financial firms lead to higher increases in house prices in the inner ring, relative to other services. These firm entries tend to be bigger in target employment and they pay higher wages. However, the effects are not statistically significant.

We also examine the effects of the high-skill firm entry on neighborhood amenities. Figure 5 summarizes the treatment effects 6–10 years later. We do not find economically large and statistically significant effects on the number of establishments per capita for restaurants, nightlife, recreation and private services.<sup>19</sup> We find weak evidence that retail amenities per capita in the inner ring decreased by 4 percent, relative to those in the outer ring.<sup>20</sup> This suggests that retail amenities like grocery stores, convenience stores, and apparel stores might be priced out of the most treated neighborhoods when commercial retail rents increase after the entry.

As a robustness check, in Figures B8 and B9, we show the estimated treatment effects on neighborhood level outcomes are robust to additionally controlling for a neighborhood’s proximity to ZIP Codes with high prior density of high-tech employment. This makes sure the identification explicitly comes from comparing neighborhoods with similar access to pre-existing high-tech jobs nearby.

### 4.3 Effects on Incumbent Outcomes

We then estimate the effects of the firm entry on incumbent residents’ consumption, finances, and mobility. Our treatment groups consists of incumbent residents living in the inner and middle rings at the time of entry announcement. The control group lives in the outer ring. Figures B11, B12 and B13 plot the event study coefficients of estimating equation (2) for each of the outcomes for years before and after the entry. Figures 6 and 7 summarize the average treatment effects for incumbent residents in the inner ring, relative to those in the outer ring, 5–8 years after the entry. The estimates are broken down by skill group and homeownership status at the time of entry announcement. All effects are presented in percent terms, so they are scaled by the mean of the control group. In our discussion of results below, we focus on comparing incumbent residents

<sup>17</sup>This is constructed as  $1 - \text{“residence own commuting share”}$  following Monte et al. (2018) who propose this as a measure of commuting openness, but adapted to the entry zone level. We use the Journey-to-Work data from 1990 and 2000 Census Transportation Planning Package (CTPP) and the Origin-Destination data from 2002-2017 LEHD Origin-Destination Employment Statistics (LODES) available prior to the firm entry announcement to construct this measure.

<sup>18</sup>Table A7 shows how we define the industry categories.

<sup>19</sup>Table A15 documents how we construct measures of these amenities.

<sup>20</sup>The effect is statistically significant at the 90% level.

living in the inner ring and in the outer ring. The treatment effects from comparing those in the middle ring and outer ring are qualitatively similar, but in general smaller in magnitudes.

Our results suggest that high-skilled incumbents, especially homeowners, benefit. We measure incumbents' non-durable consumption by the number of bank revolving accounts they own. Figure 6 shows that high-skilled owners in the inner ring increase non-durable consumption by 1.8 percent 5–8 years later, relative to those in the control group. High-skilled owners are also less likely to experience bad credit events. Those in the inner ring are 1.0 percent less likely to experience any mortgage delinquencies and 0.8 percent less likely to experience any debt collections. We do not find economically large and statistically significant effects on their durable consumption measured by whether they have any auto trades. High-skilled owners benefit from increased wages after the entry. They also receive a housing dividend due to the appreciation of their home values, relative to the renters. Hence they are able to increase consumption and have better financial health.

In contrast, high-skilled renters possibly benefit less than high-skilled owners. We do not find economically large and statistically significant effects on their credit scores, non-durable consumption, durable consumption, and bad credit events.<sup>21</sup> These results suggest high-skilled renters possibly benefit less than high-skilled owners. Their benefits from increased wages are offset by the increase in rents and increase in the prices of local non-tradable goods that are correlated with house prices.

The effects on low skilled renters are inconclusive but suggest they may benefit less or even be harmed. We do not find economically large and statistically significant effects for low-skilled renters in the most treated neighborhoods on credit score, consumption and credit events.<sup>22</sup> Since we find house prices increase more closer to the entry sites, we expect that low-skilled renters could be possibly harmed. However, among low-skilled incumbents, homeowners are better off than renters since the effects of the house price increase are more neutral to them. We find low-skilled owners in the inner ring increase non-durable consumption by 1.8 percent 5–8 years later, relative to those in the control group. We do not find economically large and statistically significant effects on their credit scores, durable consumption and credit events.<sup>23</sup>

To shed light on the current debate on whether neighborhood gentrification displaces incumbent residents, we also track incumbent residents' migration decisions. We confirm that high-skilled firm entries can induce migration among incumbent residents that looks like gentrification. We find low-skilled incumbents are more likely to be displaced further away to lower quality neighborhoods. In particular, low-skilled renters in the inner ring are 6.0 percent more likely to move more than 20 miles away from their address when the firm entry was announced, which proxies for moving out of

---

<sup>21</sup>Based on the 95% confidence intervals, we can rule out larger than 1.1 percent increases in their credit scores and durable consumption, a larger than 2.0 percent increase in non-durable consumption, and larger than 1.6 percent increase in the probability of never experiencing bad credit events.

<sup>22</sup>Based on the 95% confidence intervals, the effects on credit scores, durable consumption, and never experiencing bad credit events are between -1.4 and 1.4 percent; the effects on non-durable consumption is between -1.9 and 2.5 percent.

<sup>23</sup>Based on the 95% confidence intervals, the effects on credit scores and durable consumption are between -0.6 and 0.4 percent; the effects on never experiencing bad credit events are between -1.8 and 0.7 percent.

their original city. A caveat is that only 17 percent of the low-skilled renters in the control group move out of the city after the entry. On average, there are 8,343 incumbent low-skilled renters in the inner ring. This suggests that on average in each entry zone, the number of displaced low-skilled renters in the most treated neighborhoods is around 85 ( $0.17 \times 0.06 \times 8,343$ ) which is quite small.<sup>24</sup>

In addition, we also examine the neighborhood quality of the places incumbent residents move to after the entry. Figure 7 shows they move to neighborhoods with 1.5 percent lower income per capita and 1.5 percent lower median home value. We use the upward mobility measures from Chetty et al. (2018) as another proxy for neighborhood quality.<sup>25</sup> Low-skilled renters move to neighborhoods with lower upward mobility to the top of the income distribution. Specifically, low-skilled renters live in Census Tracts with 3.2 and 6.6 percent lower mean probability of reaching the top 1 percentile of the national income distribution, with and without conditioning on children coming from poor family with the parental income at 25 percentile of the national income distribution.<sup>26</sup> Taken together, low-skilled renters in the treatment groups on average experience a small decrease in neighborhood quality, relative to those in the control group. However, for those who get displaced, the decrease in their neighborhood quality is likely larger, given that income and house prices increase more closer to the entry.

Similarly, low-skilled owners in the inner ring are 7.6 percent more likely to move out of their original city. They are also more likely to move to neighborhoods of slightly lower quality measured by income per capita, median home value and upward mobility of children.<sup>27</sup> Compared to renters, homeowners are insured against rising rents if they stay in their original neighborhoods. However, we have shown evidence that low-skilled owners are also more likely to move further away and they increase non-durable consumption. Taken together, their wage increase after the entry could be smaller than the increase in the user cost of owning their homes and the increase in the prices of local goods related to housing. Therefore, low-skilled owners in the most treated neighborhoods are more likely to cash out from their housing assets, move to slightly cheaper neighborhoods, and increase consumption.

In summary, our reduced-form evidence suggests that high-skilled incumbents, especially homeowners, benefit. Low-skilled incumbents on average benefit less. In particular, low-skilled renters appear to be harmed.

---

<sup>24</sup>As robustness checks, in Figure B14, we also find low-skilled incumbents, both renters and owners, are more likely to move away from the address they were living at when the entry was announced. They are also more likely to move away by more than 100 miles, which proxies for moving out of the metropolitan area.

<sup>25</sup>These measures are based on the sample of children in the 1978–1983 birth cohorts who are U.S. citizen or authorized immigrants. Children’s individual incomes are measured in 2014 and 2015, when they are between the ages of 31 and 37.

<sup>26</sup>Figure B15 shows similar negative treatment effects if we condition on parental income at the 50 percentile and 75 percentile of the national income distribution. We do not find statistically significant difference in the upward mobility as measured by the probability of reaching the top quintile of the national income distribution.

<sup>27</sup>The magnitudes of the treatment effects on neighborhood quality are smaller than those for low-skilled renters, but they are significant at 95% level.



## 5 Spatial Equilibrium Model with Heterogeneous Labor

After a high-skilled firm entry, the demand for both high-skilled and low-skilled workers could change due to (1) the direct effects of the high-skilled firm entry (2) the indirect GE effects through changes in wages and rents. Our reduced-form results captures the effects due to both of channels. To tie together the reduced-form evidence into welfare implications, we estimate a model of individual home and work location choice model. The model illustrates the key economic forces that would rationalize our reduced-form evidence. We then use the model to evaluate the welfare incidence on incumbent residents by their skill group and homeownership status. In the model, workers choose a ZIP Code to live in and a ZIP Code to work in inside a city. We model workers' neighborhood choices as a function of their wages, rents, commuting cost, local amenities, moving cost, and taste heterogeneity.

### 5.1 Setup

Our model features a finite collection of neighborhoods  $\mathcal{N} = \{\mathcal{N}_1, \mathcal{N}_0\}$  within a city of 50-mile radius around the firm entry site, with  $\mathcal{N}_1$  neighborhoods inside the entry zone of all ZIP Codes within 30 mins of driving time from the entry site, and  $\mathcal{N}_0$  neighborhoods on the periphery of the city, which are ZIP Codes outside the entry zone, but within 50 miles from the entry site.<sup>28</sup> We combine neighborhoods on the periphery and the rest of the country into a single outside option, denoted as  $\mathcal{N}_\emptyset$ . We normalize the indirect utility of choosing either to live or work in the outside option to always be 0<sup>29</sup>. Workers differ in their skill level and homeownership status. There are high-skilled workers  $H$  and low-skilled workers  $L$ . They can be homeowners  $O$  or renters  $R$ .

The timings of events are as follows: in period 0, workers are born in neighborhood  $i_0$  with skill group  $s \in \{H, L\}$  and homeownership status  $o \in \{O, R\}$ . In period 1, workers choose which neighborhoods to live and work in, and the amount of consumption between a nationally traded good and a housing-related local good by maximizing individual utility. In period 2, a skill-biased productivity shock happens in a neighborhood due to a high-skilled firm entry, and workers re-optimize by choosing to where to live and work and the amount of consumption again.

### 5.2 Worker's Choice of Residential and Work Neighborhoods

A worker  $\omega$ , who belongs to skill group  $s \in \{H, L\}$  and homeownership group  $o \in \{O, R\}$ , chooses the residential neighborhood  $i \in \mathcal{N}$  and workplace neighborhood  $j \in \mathcal{N}$  in period  $t$  that gives him the most desirable bundle of wages, local good prices, commuting costs and amenities. The worker inelastically supplies one unit of labor and earns a wage of  $W_{j,t}^s$ , which differs by his skill group.

---

<sup>28</sup>In our reduced-form strategy, we compare the relative differences between neighborhoods close to and far away from the entry within the 30-min entry zone. In the model, we include the periphery of the city to allow GE effects beyond the entry zone. The 50-mile ring is chosen because it roughly corresponds to the size of a metropolitan area around the entry site, and also for computational tractability when we analyze welfare incidence for incumbents.

<sup>29</sup>This amounts to assuming the difference in the indirect utility for neighborhoods on the periphery  $\mathcal{N}_0$  and the rest of the country is the same before and after the firm entry.

The worker consumes a local good  $M$ , which is related to housing services, with a local price  $R_{i,t}$ . He also consumes a nationally traded good  $C$ , with a national price normalized to 1.

We introduce the renter maximization problem first. A worker who is a renter has Cobb-Douglas preferences for the local and national good. He maximizes the following utility function subject to his budget constraint:

$$\begin{aligned} \max_{M, C, (i, j)} \quad & \ln(M^{\theta^R}) + \ln(C^{1-\theta^R}) - \lambda^s \tau_{i,j} + a_{i,j,t}^s - b_{\omega}^s \mathbf{1}_{\omega,t} (i \neq i_0) + \sigma^s \varepsilon_{i,j,\omega,t} \\ \text{s.t.} \quad & R_{i,t} M + C \leq W_{j,t}^s. \end{aligned} \quad (3)$$

Renter's share of income spent on the local good is  $\theta^R$ , where  $0 \leq \theta^R \leq 1$ . He incurs a commuting cost of  $\lambda^s \tau_{i,j}$ , where  $\tau_{i,j}$  is the commuting time between residential neighborhood  $i$  and workplace neighborhood  $j$ , and  $\lambda^s$  is the elasticity of commuting by skill group. He gains utility from  $a_{i,j,t}^s$ , a set of exogenous local amenities that are valued differently by different skill groups<sup>30</sup>. To capture a worker's local ties to the place he initially lives in, he incurs dis-utility from not living in his "birthplace" neighborhood  $i_0$  in period 0.<sup>31</sup> The magnitude of the birthplace dis-amenity depends on worker's skill level and whether the worker is born inside or outside the entry zone. Let  $b_{\omega}^s = \sigma^s b_{\omega}$  and define  $b_{\omega} \equiv b z_{\omega}$ , where  $z_{\omega}$  is  $2 \times 1$  vector of indicator variables with each entry equal to 1 if the worker is born inside and outside the entry zone, respectively.  $b = [b_{in}, b_{out}]$  is a  $1 \times 2$  vector measuring the dis-utility of not living in one's initial location for the two groups by their birthplace. The worker gains utilities from match-specific local amenities for each pair of residential and work neighborhoods  $\sigma^s \varepsilon_{i,j,\omega,t}$ , which reflects his idiosyncratic preferences. They are drawn from an i.i.d. Type I Extreme Value (T1EV) distribution with a variance of  $\sigma^s$  that differs by skill group.

Maximization of equation (3) implies the indirect utility function for choosing residential neighborhood  $i$  and workplace neighborhood  $j$ . We normalize the indirect utility function by dividing it by  $\sigma^s$ , so that we can interpret all of the coefficients as migration elasticities. The worker who is a renter has an indirect utility over a pair of home-work locations  $(i, j)$  given by:

$$V_{i,j,\omega,t}^{s,R} = \underbrace{\beta^s (w_{j,t}^s - \theta^R r_{i,t} - \lambda^s \tau_{i,j} + a_{i,j,t}^s)}_{\delta_{i,j,t}^s: \text{Mean utilities}} - \underbrace{b_{\omega}^s \mathbf{1}_{\omega,t} (i \neq i_0)}_{\text{Birthplace disamenity}} + \underbrace{\varepsilon_{i,j,\omega,t}}_{\text{Match-specific amenities}} \quad (5)$$

where  $\beta^s \equiv 1/\sigma^s$  is the preference parameter for wages.  $w_{j,t}^s$  is the log of wage specific to his skill group  $s$  in workplace neighborhood  $j$ .  $r_{i,t}$  is log of the rent he pays in his choice of residential neighborhood  $i$ .<sup>32</sup> We use  $\delta_{i,j,t}^s$  to denote the components of the utility value of choice  $(i, j)$ , which

<sup>30</sup>We assume local amenities as exogenous as we do not see economically large and statistically significant treatment effects on amenities, as discussed in Section 4.2.

<sup>31</sup>We can think of  $i_0$  as the neighborhood that we first observe an worker along his migration history. For a worker "born" outside the entry zone in the periphery of the city, his "birthplace" is  $\mathcal{N}_0$ . When we estimate the model, an incumbent resident's "birthplace" neighborhood is set to be his residential neighborhood 5 years prior to the firm entry. This setup captures the moving costs from one's neighborhood when he is younger due to the accumulation of neighborhood-specific capital, e.g., social networks. An alternative way is to have a moving cost from leaving one's residential neighborhood in the last period, which captures similar moving dynamics.

<sup>32</sup>Note that high and low-skilled workers who live in the same neighborhood face the same housing market, so

all renters of skill group  $s$  value identically.

Next we turn to the homeowner maximization problem. In equilibrium, the cost of renting a home is about the same as the user cost of owning a home. Hence we could have one price for both renters and owners for consuming housing services. In period 1 prior to the firm entry, the homeowner owns his home in his choice of residential neighborhood. The homeowner's indirect utility function in period 1 is thus the same as that of a renter. In period 2, his home value appreciates after the firm entry. The homeowner can either stay in his original home or move away and choose his housing consumption again. We model a homeowner's decision as follows: In period 2, each owner sells the house that he owns in period 1. His capital gain should equal to the net present value of the future increases in rents at his home in period 1. The homeowner annuitizes the capital gain and receives an annual housing dividend equal to  $R_{i_1,t=2} - R_{i_1,t=1}$ .<sup>33</sup> The homeowner then chooses the amount of housing services again in period 2 by maximizing the utility function:

$$\max_{M,C,(i,j)} \ln(M^{\theta^R}) + \ln(C^{1-\theta^R}) - \lambda^s \tau_{i,j} + a_{i,j,t}^s - b_\omega \mathbf{1}_{\omega,t} (i \neq i_0) + \sigma^s \varepsilon_{i,j,\omega,t} \quad (6)$$

$$\text{s.t.} \quad R_{i,t}M + C \leq W_{j,t}^s + (R_{i_1,t=2} - R_{i_1,t=1})M_1^* \quad (7)$$

$$M_1^* = \frac{\theta^R W_{j_1,t=1}^s}{R_{i_1,t=1}} \quad (8)$$

where  $i_1$  denotes his chosen residential neighborhood in period 1, and  $j_1$  is his chosen workplace neighborhood in period 1.  $M_1^*$  is the optimal amount of housing the owner owns in period 1. After solving his maximization problem, a homeowner has an indirect utility over a pair of home-work locations  $(i, j)$  given by:

$$V_{i,j,\omega,t}^{s,O} = \begin{cases} \beta^s (w_{j,t}^s - \theta^R r_{i,t} - \lambda^s \tau_{i,j} + a_{i,j,t}^s) - b_\omega \mathbf{1}_{\omega,t} (i \neq i_0) + \varepsilon_{i,j,\omega,t} & , t = 1 \quad (9) \\ \beta^s (\ln(W_{j,t}^s + \Delta R_{i_1} M_1^*) - \theta^R r_{i,t} - \lambda^s \tau_{i,j} + a_{i,j,t}^s) - b_\omega \mathbf{1}_{\omega,t} (i \neq i_0) + \varepsilon_{i,j,\omega,t} & , t = 2 \quad (10) \end{cases}$$

where  $\Delta R_{i_1} \equiv R_{i_1,t=2} - R_{i_1,t=1}$ .

### 5.3 Labor Demand and Housing Supply

We also write down the labor demand and housing supply to complete a spatial equilibrium model. Appendix E.1–E.6 provides details for each component of the model and Appendix E.8 performs a simulation exercise. Here summarize the key details of the spatial equilibrium.

The aggregate labor supply of each workplace neighborhood is the sum of all workers who choose to work there. The labor demand model features firms in each workplace neighborhood competing in a perfectly competitive market. Each firm's production function is Cobb-Douglas with constant returns to scale. There is skill-complementary in a firm's production function in that they combine high-skilled and low-skilled labor as imperfect substitutes into production with a constant elasticity of substitution. We make the assumption that homeowners and renters face the same labor market

---

that a shock to labor demand to one group could be transmitted to another group through its effect on the housing market.

<sup>33</sup>Equivalently, the annual housing dividend equals his annual rental income if he were to rent it out.

within a given skill group, hence their wages do not differ within a given skill group.

Housing demand in each residential neighborhood is the aggregate housing demand of all workers who choose to live there. The housing supply curve in each residential neighborhood is upward sloping to capture features of a city such as zoning laws and congestion.

A spatial equilibrium is reached when both the labor and housing markets clear.

## 5.4 Estimation of Location Demand

We estimate the model on individual’s home and work location choices using a sub-sample of 129 firm entries that are announced between 2003–2010 where we are able to observe worker commuting flows between home and work neighborhoods from LODES. We set period 0 to be the years up to five years prior to the firm entry announcement. Period 1 is between  $\tau \in [-5, -1]$ , where  $\tau = 0$  is the year of entry announcement. Period 2 is between  $\tau \in [-1, 5]$ . We provide full details of how we prepare the data for estimation in Appendix C.5.

To estimate worker’s preferences for location choices, the key parameters to identify in worker’s indirect utility function are  $\beta^s, \theta^R, \lambda^s, b$ . We use a two-step estimator similar to Berry (1994) and Berry et al. (1995). In addition, we use a set of micro moments derived from individual migration history from Infutor, similar to the approach in Petrin (2002). More details of the estimation procedure are provided in Appendix E.9.

In the first step, we treat the mean utility level of each pair of home-work locations  $(i, j)$  in period  $t$  for skill group  $s$ ,  $\delta_{i,j,t}^s$ , as a parameter to estimate, in addition to estimating  $b$ . We proceed as follows. First, given a guess of  $b$ , we solve for the set of  $\delta_{i,j,t}^s$ ’s that most closely match observed and model predicted choice probabilities by skill group across all entry zones using the contraction mapping suggested by Berry et al. (1995). The mean utility values  $\delta_{i,j,t}^s$  are identified by the observed population differences across pairs  $(i, j)$  for a given skill group.<sup>34</sup>

Next, to identify  $b$ , we introduce additional micro-moments similar to the approach in Petrin (2002). To identify the parameters for birthplace disamenity, we compute model predicted moments to match the observed moments computed using individual’s migration history from Infutor. To identify  $b_{in}$ , we match the share of workers born inside the entry zone in period 0, and choose to live in a residential neighborhood  $i$  that is different from his period 0 “birthplace” neighborhood in period 1 and 2, respectively. To identify  $b_{out}$ , we match the share of workers born outside the entry zone but inside the city in period 0, and choose to live in a residential neighborhood  $i$  inside the entry zone in period 1 and 2, respectively. These moments are given by

$$\mathbb{E}[\mathbf{1}\{\omega \text{ lives in } i \neq i_0 \text{ in } t = 1 | i_0 \in \mathcal{N}_1\}], \quad (11)$$

$$\mathbb{E}[\mathbf{1}\{\omega \text{ lives in } i \in \mathcal{N}_1 \text{ in } t = 1 | i_0 \in \mathcal{N}_0\}], \quad (12)$$

$$\mathbb{E}[\mathbf{1}\{\omega \text{ lives in } i \neq i_0 \text{ in } t = 2 | i_0 \in \mathcal{N}_1\}], \quad (13)$$

$$\mathbb{E}[\mathbf{1}\{\omega \text{ lives in } i \in \mathcal{N}_1 \text{ in } t = 2 | i_0 \in \mathcal{N}_0\}]. \quad (14)$$

---

<sup>34</sup>In absence of the birthplace disamenity which varies at the individual level, the estimated mean utility values for each pair of locations  $(i, j)$  would exactly equal the log population by skill group choosing each pair.

In the second step of estimation, we decompose the estimated mean utility values into how workers value wages, rents, and amenities. Taking the difference of the mean utility estimates across period 1 and 2:

$$\Delta\delta_{i,j}^s = \beta^s (\Delta w_j^s - \theta^R \Delta r_i + \Delta a_{i,j}^s) \quad (15)$$

$$= \beta^s (\Delta w_j^s - \theta^R \Delta r_i) + \beta_0^s + \Delta\xi_{i,j}^s. \quad (16)$$

Note the commuting cost term drops out. We include a constant,  $\beta_0^s$ , to allow the relative difference in utility between choices inside and outside the entry zone to change across the two periods for each skill group  $s$ . As econometricians, we observe changes in wages in workplace neighborhoods,  $\Delta w_j^s$ , and changes of rents in residential neighborhoods,  $\Delta r_i$ . However, we do not observe  $\Delta\xi_{i,j}^s \equiv \beta^s \Delta a_{i,j}^s$ , which is the change in exogenous amenities across the two periods for the pair  $(i, j)$  for workers of skill group  $s$ .

The parameter  $\theta^R$  represents worker's expenditure share on local goods. We calibrate  $\theta^R$  to be 0.62, following [Moretti \(2013\)](#) and [Diamond \(2016\)](#). To identify  $\beta^s$ , we need to use variation in changes in wages and/or rents across workplace and residential neighborhood choices that are uncorrelated with unobserved changes in local amenities. We can use the 1990 level driving time from residential neighborhood  $i$  to firm's entry site as an instrument for changes in worker's real wages  $\Delta w_j^s - \theta^R \Delta r_i$  for each pair  $(i, j)$ . Driven by the labor demand shock, high-skilled workers would want to live closer to the firm entry site to take advantage of the higher wages in nearby workplaces. Therefore, they will drive up the rents more in residential neighborhoods closer to firm's entry site. The identification assumption is that the driving time instrument provides variation in changes in rents after the firm entry, hence variation in changes in real wages, that is uncorrelated with changes in local exogenous amenities,  $\Delta\xi_{i,j}^s$ .

A threat to identification is that neighborhoods close to and far away from the actual entry site could have differences that correlate with the likelihood of attracting a firm entry nearby. For example, they could differ in their proximity to entry sites with amenities that more desirable for high-skilled firms. To address this issue, we further control for a residential neighborhood's proximity to nearby potential sites for entry, conditional on the workplace neighborhood and the entry zone. Specifically, we add a set of fixed effects, denoted by  $\phi_{h^1, h^2, j}^s$ , to equation (16). These are distance bins from residential neighborhood  $i$  to nearby potential sites for high-skilled firm entry, that we used in our reduced-form strategy in Section 3.2, interacted with the entry zone  $k$  and the workplace neighborhood  $j$ .<sup>35</sup> Therefore, we have the following final second-step estimating equation:

$$\Delta\delta_{i,j}^s = \beta^s (\Delta w_j^s - \theta^R \Delta r_i) + \phi_{h^1, h^2, j}^s + \Delta\xi_{i,j}^s. \quad (17)$$

---

<sup>35</sup>For brevity in our notations, we omit the subscript  $k$ , which denotes the entry zone. In practice, a pair of locations  $(i, j)$  could be within the 50-mile radius of more than one firm entry site. Further, note that  $\phi_{h^1, h^2, j}^s$  subsumes the constant  $\beta_0^s$  in equation (16).

The full identification assumption is the following. If we condition on the workplace neighborhood in an entry zone, the changes in wages are fixed; If we further control for a residential neighborhood's proximity to nearby potential sites for entry, its driving time to the actual entry site is correlated with the changes in rents in the residential neighborhood, but is uncorrelated with unobserved changes in exogenous local amenities. Specifically, the moment restrictions are:

$$\mathbb{E} [\Delta \xi_{i,j}^s \cdot \mathbf{1}\{t > 0\} \cdot \text{Driving Time}_i] = 0, \forall i, j \quad (18)$$

where  $\mathbf{1}\{t > 0\}$  is an indicator for time periods 1 and 2 for which we take the difference in the neighborhood amenities  $\Delta \xi_{i,j}^s$ .  $\text{Driving Time}_i$  is the driving time from residential neighborhood  $i$  to the firm's entry site.

Last, to identify the semi-elasticity of commuting, we estimate a gravity equation that regresses log of market shares by skill on driving time between residential neighborhood  $i$  and workplace neighborhood  $j$ , with fixed effects for  $i$ ,  $j$  and the entry zone  $k$ .

$$\ln (\pi_{i,j,k,t}^s) = \alpha_i + \rho_j + \eta_k - (\beta^s \lambda^s) \tau_{i,j} + \epsilon_{i,j,t}. \quad (19)$$

The derivation of equation (19) is provided in Appendix E.9.1.

After first calibrating  $\theta^R$  and estimating  $\beta^s \lambda^s$ , we jointly estimate  $b_{in}$ ,  $b_{out}$ ,  $\beta^H$ ,  $\beta^L$  via 2-step GMM. Full details of estimation are provided in Appendix E.9.

## 5.5 Parameter Estimates

Table 4 shows the GMM estimates of model parameters. Both high and low-skilled workers prefer wages, lower rents and dislike commuting. High-skilled workers have a migration elasticity with respect to real wages of  $\beta^H = 1.34$ , while low-skilled workers are slightly less responsive, with an elasticity of  $\beta^L = 1.32$ . The low values of  $\beta^s$  imply workers have heterogeneous idiosyncratic preferences for locations. Our estimates of  $\beta^s$  are in line with the extensive-margin labor supply elasticities found in studies of local labor markets. For example, [Severen \(2018\)](#) finds a labor supply elasticity of around 1.83, [Falch \(2010\)](#) finds it between 1.0 and 1.9, [Suárez Serrato and Zidar \(2016\)](#) find it between 0.75 and 4.2, and [Albouy and Stuart \(2020\)](#) find a value of 1.98. The semi-elasticities of commuting  $\beta^s \lambda^s$  are found to be around 0.02 for both high and low-skilled workers, which is in line with previous estimates from [Tsivanidis \(2019\)](#), who finds it to be 0.012, [Ahlfeldt et al. \(2015\)](#), who finds it to be 0.01, and [Gorback \(2020\)](#), who finds it between 0.01 and 0.02. Workers' preference to live in their initial locations 5 years prior to the firm entry are captured by the birthplace disamenity estimates. High-skilled workers initially live inside the entry zone at  $\tau = -5$  are 5.1 times more likely to live in their "birthplace" ZIP Code than other ZIP Codes; High-skilled workers initially live in the periphery of the city are 7.5 more likely to live in the periphery than other places. The preferences for local ties for low-skilled workers are similar.

## 5.6 Welfare Analysis

We use the model estimates to calculate the welfare incidence of high-skilled firm entries on local incumbent residents who live and work within 50 miles of firm's entry site in period 1 prior to the entry. We then compare the expected welfare change for incumbent workers by their skill and homeownership status at the time of firm entry. We further compare the the expected welfare change for those who stayed in their original neighborhood with those move away to a different neighborhood.

We can write the expected utility of workers from demographic group  $\{s, o\}$  in period  $t$  as:

$$\mathbb{E}[U_t^{s,o}] = \mathbb{E}_\varepsilon \left[ \max_{i,j} \{v_{i,j,\omega,t}^{s,o} + \sigma^s \varepsilon_{i,j,\omega,t}\} \right] \quad (20)$$

$$= \sigma_s \sum_{i_0 \in \mathcal{N}} \pi_{i_0}^{s,o} \ln \left( 1 + \sum_i \sum_j \exp \left( \frac{v_{i,j,\omega,t}^{s,o,i_0}}{\sigma_s} \right) \right), \quad (21)$$

where  $\pi_{i_0}^{s,o}$  denotes the share of workers from group  $\{s, o\}$  born in neighborhood  $i_0$ . The expectation operator  $\mathbb{E}_\varepsilon$  is defined over the idiosyncratic preference term  $\varepsilon_{i,j,\omega,t}$ . Equation (21) is derived from the fact that  $\varepsilon_{i,j,\omega,t}$  is distributed T1EV.  $v_{i,j,\omega,t}^{s,o,i_0}$  denotes the mean utility of the pair  $(i, j)$  in period  $t$  for workers of type  $\{s, o, i_0\}$ .

We first compute the expected utility in period 1 using the observables in period 1, which is one year prior to the firm entry announcement. Next, we compute the expected utility in period 2 by adjusting wages and rents by the amount of change due to the firm entry itself, while fixing the commuting costs and local amenities at their period 1 level. To estimate the changes in wages and rents for neighborhoods within each entry zone, which are due to the firm entry alone, we estimate the following equations:

$$\Delta w_{j,k}^s = \beta_d \text{Driving Time}_{j,k} + \psi_{h^1,h^2,j,k}^s + \epsilon_{j,k}, \quad (22)$$

$$\Delta r_{i,k} = \gamma_d \text{Driving Time}_{i,k} + \phi_{h^1,h^2,i,k} + \epsilon_{i,k}, \quad (23)$$

where  $\psi_{h^1,h^2,j,k}^s$  and  $\phi_{h^1,h^2,i,k}$  are the same fixed effects we defined in Section 3.2 which control for a workplace neighborhood and a residential neighborhood's proximity to nearby potential sites for entry, respectively;  $\text{Driving Time}_{j,k}$  and  $\text{Driving Time}_{i,k}$  are driving time from each workplace and residential neighborhood to the entry site  $k$ , respectively. To compute the expected utility in period 2, we let the wages and rents change by the amount implied by the firm entry from equations (22) and (23):  $\Delta \hat{w}_{j,k}^s = \hat{\beta}_d \text{Driving Time}_{j,k}$  and  $\Delta \hat{r}_{i,k} = \hat{\gamma}_d \text{Driving Time}_{i,k}$ .

We then compute the change in expected utility between period 1 and 2:

$$\mathbb{E}[\Delta U^{s,o}] = \mathbb{E}[\hat{U}_{t=2}^{s,o}] - \mathbb{E}[U_{t=1}^{s,o}] \quad (24)$$

Recall that  $\sigma^s = 1/\beta^s$  in equation (5), where  $\beta^s$  is the marginal utility of log earnings. Hence,



$\mathbb{E}[\Delta U^{s,o}]$  is measured in the unit of log earnings.<sup>36</sup> Finally, we multiply  $\mathbb{E}[\Delta U^{s,o}]$  by worker’s expected wage in period 1 measured in 2010 dollars to convert welfare changes into dollar units.

Our proposed method to evaluate the welfare incidence of a high-skilled firm entry captures both the direct effects of the firm entry through the jobs it created, as well as the indirect GE effects through changes in house prices and wages that are correlated with the distance to the firm entry shock. However, we do not capture the city-wide GE effects. This is a similar limitation as our reduced-form strategy which compares the relative differences between neighborhoods close to and far away. An alternative strategy that allows us to capture the full GE effects including the city-wide GE effects is to estimate the rest of the spatial equilibrium model including labor demand and housing supply and demand. The trade-offs are additional functional assumptions about firm’s production function and housing supply, as well as additional identification challenges for estimation.

We compute the change in expected utility for incumbent residents who live and work within 50 miles from the entry site in period 1. First, we assume they are fixed at his period 1 choices of both residential and workplace neighborhoods, but they are allowed to re-optimize his consumption of the national good and the local goods. We fix commuting cost and local amenities at their period 1 level. For each of 129 firm entries in our estimation sample, we decompose the expected welfare change due to the firm entry by the components of worker’s utility function, and report the average welfare change across the firm entries. These welfare changes can be interpreted as the welfare change for a single representative firm entry in our sample.

Panel A of Table 5 reports welfare changes for this “no-sorting” scenario. Because no sorting happens in period 2, there are no welfare changes due to commuting, local amenities, birthplace disamenities and match-specific amenities, grouped into the category of “Others”. For a representative firm entry, homeowners benefit. In particular, high-skilled homeowners receive an annual benefit equivalent to \$110 in 2010 real dollars, five years after the entry. Correspondingly, they receive an annual welfare increase equivalent to a 0.21% increase in their average wages in period 1. In contrast, low-skilled owners have a much smaller welfare increase of \$21 annually (0.07% increase in their period 1 wages). We find that high-skilled owners benefit the most. High-skilled wages increase much more than low-skilled, as shown in Figure B16a. Compared to the renters, owners do not face the increasing rents that renters face; So they are better off in relative terms. In our model, homeowners receive an annual housing dividend from the homes they own in period 1. Even though rents increase in period 2, with this housing dividend, homeowners would be strictly better off by decreasing their housing consumption a little bit and consume more of the national good.<sup>37</sup>

---

<sup>36</sup>For example, if  $\mathbb{E}[\Delta U^{s,o}]$  equals to 0.1, then welfare change is equivalent to a 10% increase in a worker’s wage in period 1.

<sup>37</sup>An intuitive proof works as follows: In period 2, suppose wages are fixed at their period 1 level, and owners are not allowed to adjust housing consumption. When rents increase after the entry, owners are perfectly insured and there should be no welfare changes for them. If we allow owners to adjust housing consumption, because the implicit rental costs have increased relative to the price of the national good, price effect predicts that an owner should substitute some housing consumption with consumption of the national good, and thus be strictly better-off

On the other hand, renters are hurt. In particular, high-skilled renters have a welfare decrease of \$32 annually (0.06% decrease in their period 1 wages). Low-skilled renters are hurt the most with a welfare decrease of \$55 annually (0.2% decrease in their period 1 wages). High-skilled renters are hurt slightly because the increase in their wages are more than off-set by the increase in rents everywhere. Low-skilled renters are especially harmed because their wages increase much less than the increase in rents. Figure B16c shows that for most of the home-work locations,  $(i, j)$ , the changes in the mean utilities for low-skilled renters,  $\Delta\delta_{i,j}^L$ , after the entry that are due to changes in real wages  $(\Delta w_j^L - \theta^R \Delta r_i)$  are negative.

In Panel B of Table 5, we allow workers to re-optimize by choosing the residential and workplace neighborhoods that maximize their utility in response to the firm entry, and decompose their welfare changes. We have the same patterns in welfare changes as in the “no-sorting” scenario. Homeowners benefit, and high-skilled owners benefit the most. Renters are hurt, and low-skilled renters are hurt the most. Across all four demographic groups, the welfare changes only slightly differ from those in the case of not allowing sorting in period 2. This is because welfare changes are to first order driven by changes in prices such as changes in wages and rents, which can be shown using the Envelope Theorem.<sup>38</sup> Welfare changes due to workers re-sorting are of second order in comparison to price changes.

To understand the patterns in welfare changes in the “sorting” scenario, in Figures B17 and B18, we calculate the model predicted changes in choice probabilities of different residential and workplace neighborhoods after the entry. We take high-skilled owners who benefit the most for an illustration. They are able to work closer to the firm entry site to take advantage of the higher wage increases there. They also live further away from the entry site to avoid the higher increase in rents closer to the entry. Because rents increase everywhere and more high-skilled owners move in from outside the entry zone, the overall increase in rental costs more than offset the housing dividends from the homes they own in period 1. Longer commuting trips and living further away from their initial locations, grouped into the “Others” category, lower the overall welfare benefits for high-skilled homeowners. On the other hand, low-skilled renters are hurt the most because rents increase much more than the increase in low-skilled wages. Low-skilled renters are forced to live in cheaper neighborhoods further away from the entry or even move out of the entry zone. Consequently, they work in neighborhoods further away from the entry with lower wages.

Overall, the average welfare change for an incumbent living in the metropolitan area around the firm’s entry location seems small. For example, the average welfare change for low-skilled renters is equivalent to an annual 0.2 percent decrease in their initial wages. This change is much smaller than the earning loss due to a typical job displacement. For example, Couch and Placzek (2010) find the earning loss of a job displacement is 15 percent six years later for workers in Connecticut. The welfare change per incumbent is small suggesting that not everyone in the metropolitan area will be so affected by the firm entry. Table A13 shows the welfare changes are heterogeneous by where

---

in his utility. With increases in wages in period 2, owners would be even more better-off.

<sup>38</sup> Appendix E.7 provides a proof.

incumbent residents are initially located five years prior to the entry. In general, the magnitudes of the welfare changes are bigger the closer they are to entry site initially, as both wages and rents increased more closer to the entry site.

However, the aggregate welfare changes across all incumbent residents within 50 miles of the firm entry are substantial. In particular, high-skilled incumbents benefit by an annual \$23.2 million; Low-skilled incumbents lose by an annual \$13.2 million. Overall, the annual aggregate welfare change for a representative firm entry in our sample with around 1,000 promised jobs amounts to \$10 million. The overall positive welfare change is driven by the benefits of homeowners who have a large share among the incumbents. However, renters’ welfare are harmed. In comparison, the average one-time discretionary subsidy for a firm entry is \$107 million for 1,000 promised jobs (Slattery, 2019).<sup>39</sup> Hence it would take more than 10 years for the benefits to justify the firm entry subsidies.

An caveat is that the welfare changes we calculate are for a single representative firm entry. These effects could potentially be larger with multiple firm entries into an area within a short period of time.

### 5.6.1 Mover vs. Stayers

Local governments are often concerned about the potential negative welfare effects of high-skilled firm entry on local incumbent residents, especially on those who are displaced as a result. We therefore do a comparison of the welfare incidence for incumbents who stay in vs. move away from their original neighborhood in Table A14. We define stayers as incumbent residents who stayed in their period 0 “birthplace” residential neighborhood in both period 1 and period 2; movers are those who stay in their period 0 “birthplace” residential neighborhood in period 1, and move to a different residential neighborhood in period 2<sup>40</sup>. Appendix E.10 provides a more detailed discussion. Overall, we find a small difference in welfare changes between the movers and stayers. We also find similar patterns of welfare change among incumbents of different skill and homeownership status within the subgroups of movers and stayers: homeowners benefit and renters lose, and high-skilled homeowners gains the most and low-skilled renters are hurt the most. Further, low-skilled renters among the movers are hurt more than those among the stayers, which are consistent with renters who left their original neighborhoods are more likely to be initially in neighborhoods that receive higher increase in rents due to the firm entry, and therefore their welfare loss is bigger.

## 6 Conclusion

In this paper, we study the aggregate and distributional consequences of high-skilled firm entries on incumbent residents. To study this, we construct a dataset of 391 such entries in the U.S.

<sup>39</sup>Slattery (2019) collects a dataset on firm specific subsidies and reports that the average discretionary subsidy for a firm entry is \$160 million for 1,500 promised jobs. We scale the number accordingly.

<sup>40</sup>For incumbent residents born outside the entry zone, stayers are those who live outside the entry zone in both periods; movers are those who live outside in period 1, and live inside the entry zone in period 2.

from 1990–2010. We bring to bear novel rich micro-data on individual address histories, property characteristics, and financial records. Using these data, we track incumbent residents’ outcomes over 13 years before and after the firm entry and study how they are differentially affected by the entry depending on their skill and homeownership status. To do so, we use a reduced-form strategy where we control for a neighborhood’s proximity to nearby potential sites for high-skilled firm entry. We use the remaining variation in the distance from incumbent residents’ neighborhood to where the firm enters to identify the effects of the entry on incumbent residents due to their differential exposure to the firm entry shock. We estimate the effects of the firm entry on incumbent residents’ consumption, finances, and mobility. We then estimate a model of individual home and work location choices to quantify welfare incidence by incumbent residents’ skill and homeownership status.

Taken together, our results show high-skilled incumbents, especially homeowners, benefit. Low-skilled owners benefit less than high-skilled owners. Low-skilled renters are harmed. Across all the incumbents in the metropolitan area around the firm entry site, the welfare change per capita is small. After five years, low-skilled renters incur an annual welfare loss that is equivalent to a 0.2 percent decline in their initial wages one year prior to the entry. The welfare changes are heterogeneous by incumbent residents’ initial locations, with those who lived close to the entry disproportionately affected. Further, renters who move away from their original neighborhoods are hurt more than renters who stayed.

However, the aggregate welfare changes across all incumbents in the metropolitan area for a representative firm entry are quite substantial. On aggregate, high-skilled incumbents benefit by an annual \$23.2 million and low-skilled incumbents lose by an annual \$13.2 million. Benefits accrue to the homeowners at the cost of renters.

These results highlight the distributional consequences to be considered when local government compete for large firm entries with financial incentives like firm-specific subsidies and tax incentives. A point of future research would be to design an optimal housing assistance program to help incumbent residents who are disproportionately affected by the firm entry.

## References

- AHLFELDT, G. M., S. J. REDDING, D. M. STURM, AND N. WOLF (2015): “The Economics of Density: Evidence From the Berlin Wall,” *Econometrica*, 83, 2127–2189.
- ALBOUY, D. AND B. A. STUART (2020): “Urban Population and Amenities: The Neoclassical Model of Location,” *International Economic Review*, 61, 127–158.
- ALMAGRO, M. AND T. DOMINGUEZ-IINO (2020): “Location Sorting and Endogenous Amenities: Evidence from Amsterdam,” *Working paper*, 93.
- AUTOR, D. H., D. DORN, AND G. H. HANSON (2013): “The China Syndrome: Local Labor Market Effects of Import Competition in the United States,” *American Economic Review*, 103, 2121–2168.
- BARTIK, A. W., J. CURRIE, M. GREENSTONE, AND C. R. KNITTEL (2019): “The Local Economic and Welfare Consequences of Hydraulic Fracturing,” *American Economic Journal: Applied Economics*, 11, 105–155.
- BARTIK, T. (2018): “Who Benefits From Economic Development Incentives? How Incentive Effects on Local Incomes and the Income Distribution Vary with Different Assumptions about Incentive Policy and the Local Economy,” *Upjohn Institute Technical Reports*.
- BARTIK, T. J. (2017): “A New Panel Database on Business Incentives for Economic Development Offered by State and Local Governments in the United States,” *Upjohn Institute Technical Reports*, 131.
- BAUM-SNOW, N., D. HARTLEY, AND K. OK LEE (2019): “The Long-Run Effects of Neighborhood Change on Incumbent Families,” Tech. rep., Federal Reserve Bank of Chicago.
- BERNSTEIN, S., R. DIAMOND, T. MCQUADE, AND B. POUSADA (2019): “The Contribution of High-Skilled Immigrants to Innovation in the United States,” *Working paper*, 73.
- BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): “Automobile Prices in Market Equilibrium,” *Econometrica*, 63, 841–890.
- BERRY, S. T. (1994): “Estimating Discrete-Choice Models of Product Differentiation,” *The RAND Journal of Economics*, 25, 242–262.
- BLATTNER, L. AND S. NELSON (2020): “How Costly is Noise? Data and Disparities in the US Mortgage Market,” *Working paper*.
- BRUMMET, Q. AND D. REED (2019): “The Effects of Gentrification on the Well-Being and Opportunity of Original Resident Adults and Children,” Working Paper (Federal Reserve Bank of Philadelphia) 19-30, Federal Reserve Bank of Philadelphia.
- BUSO, M., J. GREGORY, AND P. KLINE (2013): “Assessing the Incidence and Efficiency of a Prominent Place Based Policy,” *American Economic Review*, 103, 897–947.
- CALDER-WANG, S. (2019): “The Distributional Impact of the Sharing Economy on the Housing Market,” *Working paper*, 84.

- CHETTY, R., J. N. FRIEDMAN, N. HENDREN, M. R. JONES, AND S. R. PORTER (2018): “The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility,” *NBER Working Paper NO. 25147*, 94.
- CHETVERIKOV, D., B. LARSEN, AND C. PALMER (2016): “IV Quantile Regression for Group-Level Treatments, With an Application to the Distributional Effects of Trade,” *Econometrica*, 84, 809–833.
- CONSUMER FINANCIAL PROTECTION BUREAU (2014): “Using Publicly Available Information to Proxy for Unidentified Race and Ethnicity: A Methodology and Assessment,” [https://files.consumerfinance.gov/f/201409\\_cfpb\\_report\\_proxy-methodology.pdf](https://files.consumerfinance.gov/f/201409_cfpb_report_proxy-methodology.pdf).
- CONWAY (2018): “Conway Analytics,” <https://conway.com/analytics/>.
- COUCH, K. A. AND D. W. PLACZEK (2010): “Earnings Losses of Displaced Workers Revisited,” *American Economic Review*, 100, 572–589.
- COUTURE, V. (2016): “Valuing the Consumption Benefits of Urban Density,” *Working paper*, 52.
- COUTURE, V., C. GAUBERT, J. HANDBURY, AND E. HURST (2019): “Income Growth and the Distributional Effects of Urban Spatial Sorting,” *Working paper*, 87.
- COUTURE, V. AND J. HANDBURY (2019): “Urban Revival in America, 2000 to 2010,” *NBER Working Paper*.
- CURRIE, J., L. DAVIS, M. GREENSTONE, AND R. WALKER (2015): “Environmental Health Risks and Housing Values: Evidence from 1,600 Toxic Plant Openings and Closings,” *American Economic Review*, 105, 678–709.
- DIAMOND, R. (2016): “The Determinants and Welfare Implications of US Workers’ Diverging Location Choices by Skill: 1980-2000,” *American Economic Review*, 106, 479–524.
- DIAMOND, R., A. GUREN, AND R. TAN (2020): “The Effect of Foreclosures on Homeowners, Tenants, and Landlords,” Tech. Rep. w27358, National Bureau of Economic Research, Cambridge, MA.
- DIAMOND, R. AND T. MCQUADE (2019): “Who Wants Affordable Housing in Their Backyard? An Equilibrium Analysis of Low-Income Property Development,” *Journal of Political Economy*, 127, 1063–1117.
- DIAMOND, R., T. MCQUADE, AND F. QIAN (2019): “The Effects of Rent Control Expansion on Tenants, Landlords, and Inequality: Evidence from San Francisco,” *American Economic Review*, 109, 3365–3394.
- DING, L., J. HWANG, AND E. DIVRINGI (2016): “Gentrification and Residential Mobility in Philadelphia,” *Working paper*, 50.
- DRAGAN, K., I. G. ELLEN, AND S. GLIED (2020): “Does Gentrification Displace Poor Children and Their Families? New Evidence from Medicaid Data in New York City,” *Regional Science and Urban Economics*, 83, 103481.
- ECKERT, F., S. GANAPATI, AND C. WALSH (2020): “Skilled Scalable Services: The New Urban Bias in Economic Growth,” *Working paper*, 108.

- EINSTEIN, K., D. GLICK, M. PALMER, AND S. FOX (2018): “Menino Survey of Mayors 2018,” Tech. rep., Boston University Initiative on Cities.
- ELLEN, I. G. AND K. M. O’REGAN (2011): “How Low Income Neighborhoods Change: Entry, Exit, and Enhancement,” *Regional Science and Urban Economics*, 41, 89–97.
- ELLIOTT, M. N., P. A. MORRISON, A. FREMONT, D. F. MCCAFFREY, P. PANTOJA, AND N. LURIE (2009): “Using the Census Bureau’s Surname List to Improve Estimates of Race/Ethnicity and Associated Disparities,” *Health Services and Outcomes Research Methodology*, 9, 69–83.
- FAJGELBAUM, P. D., E. MORALES, J. C. SUÁREZ SERRATO, AND O. ZIDAR (2019): “State Taxes and Spatial Misallocation,” *The Review of Economic Studies*, 86, 333–376.
- FALCH, T. (2010): “The Elasticity of Labor Supply at the Establishment Level,” *Journal of Labor Economics*, 28, 237–266.
- FEYRER, J., E. T. MANSUR, AND B. SACERDOTE (2017): “Geographic Dispersion of Economic Shocks: Evidence from the Fracking Revolution,” *American Economic Review*, 107, 1313–1334.
- FREEMAN, L. (2005): “Displacement or Succession?: Residential Mobility in Gentrifying Neighborhoods,” *Urban Affairs Review*, 40, 463–491.
- FREEMAN, L. AND F. BRACONI (2004): “Gentrification and Displacement New York City in the 1990s,” *Journal of The American Planning Association*, 70, 39–52.
- GIROUD, X. AND J. RAUH (2019): “State Taxation and the Reallocation of Business Activity: Evidence from Establishment-Level Data,” *Journal of Political Economy*, 127, 1262–1316.
- GOODMAN, J. D. (2018): “Amazon Is Getting at Least \$1.7 Billion to Come to Queens. Now Comes the Fight Over Whether It’s Worth It.” *The New York Times*.
- GORBACK, C. (2020): “Your Uber Has Arrived: Ridesharing and the Redistribution of Economic Activity,” *Working paper*, 78.
- GREENSTONE, M., R. HORNBECK, AND E. MORETTI (2010): “Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings,” *Journal of Political Economy*, 118, 536–598.
- HANSEN, L. P. (1982): “Large Sample Properties of Generalized Method of Moments Estimators,” *Econometrica*, 50, 1029–1054.
- HOLIAN, M. J. AND M. E. KAHN (2015): “Household demand for low carbon policies: Evidence from California,” *Journal of the Association of Environmental and Resource Economists*, 2, 205–234.
- HORNBECK, R. AND E. MORETTI (2018): “Who Benefits From Productivity Growth? Direct and Indirect Effects of Local TFP Growth on Wages, Rents, and Inequality,” Tech. Rep. w24661, National Bureau of Economic Research, Cambridge, MA.
- MARTIN, I. W. AND K. BECK (2018): “Gentrification, Property Tax Limitation, and Displacement,” *Urban Affairs Review*, 54, 33–73.



- McFADDEN, D. (1973): “Conditional Logit Analysis of Qualitative Choice Behavior,” in *Frontiers in Econometrics*, ed. by P. Zarembka, New York: Academic Press, 105–42.
- McKINNISH, T., R. WALSH, AND T. KIRK WHITE (2010): “Who Gentrifies Low-Income Neighborhoods?” *Journal of Urban Economics*, 67, 180–193.
- MONTE, F., S. J. REDDING, AND E. ROSSI-HANSBERG (2018): “Commuting, Migration, and Local Employment Elasticities,” *American Economic Review*, 108, 3855–3890.
- MORETTI, E. (2013): “Real Wage Inequality,” *American Economic Journal: Applied Economics*, 5, 65–103.
- MUEHLENBACHS, L., E. SPILLER, AND C. TIMMINS (2015): “The Housing Market Impacts of Shale Gas Development,” *American Economic Review*, 105, 3633–3659.
- NATIONAL SCIENCE FOUNDATION (2020): “S&E Indicators 2014 - NAICS Codes That Constitute High-Technology Industries,” <https://nsf.gov/statistics/seind14/index.cfm/chapter-8/tt08-a.htm>. Accessed April 13, 2020.
- PETRIN, A. (2002): “Quantifying the Benefits of New Products: The Case of the Minivan,” *journal of political economy*, 25.
- IPUMS NHGIS (2020): <https://www.nhgis.org/>.
- ROBACK, J. (1982): “Wages, Rents, and the Quality of Life,” *Journal of Political Economy*, 90, 1257–1278.
- ROSEN, S. (1979): “Wage-Based Indexes of Urban, Quality of Life,” *Current issues i urban economics. - Baltimore u.a. : Johns Hopkins Univ. Pr., ISBN 0801821096. - 1979, p. 74-104.*
- SAIZ, A. (2010): “The Geographic Determinants of Housing Supply\*,” *The Quarterly Journal of Economics*, 125, 1253–1296.
- SEVEREN, C. (2018): “Commuting, Labor, and Housing Market Effects of Mass Transportation: Welfare and Identification,” Working Paper (Federal Reserve Bank of Philadelphia) 18-14, Federal Reserve Bank of Philadelphia.
- SLATTERY, C. AND O. ZIDAR (2020): “Evaluating State and Local Business Incentives,” *Journal of Economic Perspectives*, 34, 90–118.
- SLATTERY, C. R. (2019): “Bidding for Firms: Subsidy Competition in the U.S.” *Working paper*, 82.
- SPATIAL STRUCTURES IN THE SOCIAL SCIENCES (2020): “Diversity and Disparities,” <https://s4.ad.brown.edu/Projects/Diversity/Researcher/LTDB1.htm>.
- SU, Y. (2019): “The Rising Value of Time and the Origin of Urban Gentrification,” *Working paper*.
- SUÁREZ SERRATO, J. C. AND O. ZIDAR (2016): “Who Benefits from State Corporate Tax Cuts? A Local Labor Markets Approach with Heterogeneous Firms,” *American Economic Review*, 106, 2582–2624.
- TRAIN, K. (2009): *Discrete Choice Methods with Simulation*, Cambridge University Press.

- TSIVANIDIS, N. (2019): “Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogotá’s TransMilenio,” *Working paper*, 114.
- VIGDOR, J. L. (2002): “Does Gentrification Harm the Poor?” *Brookings-Wharton Papers on Urban Affairs*, 133–182.
- YAGAN, D. (2019): “Employment Hysteresis from the Great Recession,” *Journal of Political Economy*, 127, 2505–2558.

Table 1: Summary Statistics for Sample of Firms

|                              | Mean   | S.D.   | P10    | P90    |
|------------------------------|--------|--------|--------|--------|
| Distance to CBSA Center (km) | 22.6   | 18.3   | 4.1    | 47.4   |
| Jobs Promised                | 1,091  | 932    | 500    | 2,000  |
| Investment (USD million)     | 245    | 540    | 8      | 700    |
| Median Household Income      | 60,417 | 27,002 | 30,985 | 99,237 |
| College Share                | 0.25   | 0.17   | 0.07   | 0.48   |
| Renter Share                 | 0.37   | 0.24   | 0.11   | 0.74   |
| Observations                 | 391    |        |        |        |

Notes: Distance to CBSA Center is a proxy for the distance to downtown. Jobs Promised and Investment are number of jobs promised and investment amount at the time the firm entry was announced. Median Household Income, College Share, and Renter Share are based on 2010 Census Tracts of the firm entry. Mean and S.D. are the mean and standard deviation across the 391 firm entries. P10 and P90 are the 10th and 90th percentile values.

Table 2: Summary Statistics for Sample of Individuals and Neighborhoods

|                                             | All<br>(1)           | Inner Ring<br>(2)    | Outer Ring<br>(3)    | Difference<br>(4)  |
|---------------------------------------------|----------------------|----------------------|----------------------|--------------------|
| <b>Panel A Individual Outcomes</b>          |                      |                      |                      |                    |
| <b>Panel A.1. Demographics</b>              |                      |                      |                      |                    |
| Age at $\tau=0$                             | 43.646<br>(10.382)   | 43.547<br>(10.398)   | 43.655<br>(10.384)   | -0.107<br>(0.004)  |
| Tenure at $\tau=0$                          | 5.525<br>(5.664)     | 5.623<br>(5.740)     | 5.483<br>(5.646)     | 0.140<br>(0.002)   |
| College Share (High-skilled Share)          | 0.359<br>(0.480)     | 0.408<br>(0.491)     | 0.339<br>(0.473)     | 0.069<br>(0.0002)  |
| Share Homeowner at $\tau=0$                 | 0.531<br>(0.499)     | 0.496<br>(0.500)     | 0.542<br>(0.498)     | -0.046<br>(0.0002) |
| <i>Number of Residents Per Entry</i>        | 310,142              | 25,281               | 157,900              | 183,181            |
| <b>Panel A.2. Migration Outcomes</b>        |                      |                      |                      |                    |
| Lived $\geq 20$ miles at $\tau = -4$        | 0.088<br>(0.284)     | 0.091<br>(0.282)     | 0.089<br>(0.288)     | 0.002<br>(0.0001)  |
| <i>Number of Residents Per Entry</i>        | 310,142              | 25,281               | 157,900              | 183,181            |
| <b>Panel A.3. Financial Outcomes</b>        |                      |                      |                      |                    |
| Credit Score                                | 682.243<br>(116.830) | 682.461<br>(117.812) | 683.112<br>(115.819) | -0.651<br>(0.084)  |
| Bank Revolving Accounts                     | 3.255<br>(2.967)     | 3.205<br>(2.935)     | 3.280<br>(2.972)     | -0.075<br>(0.002)  |
| Has Auto Trades                             | 0.760<br>(0.427)     | 0.740<br>(0.439)     | 0.770<br>(0.421)     | -0.030<br>(0.0003) |
| No Mortgage Delinquencies                   | 0.454<br>(0.498)     | 0.465<br>(0.499)     | 0.454<br>(0.498)     | 0.011<br>(0.0004)  |
| No Debt Collections                         | 0.622<br>(0.485)     | 0.625<br>(0.484)     | 0.623<br>(0.485)     | 0.002<br>(0.0004)  |
| <i>Number of Residents Per Entry</i>        | 54,200               | 4,421                | 27,644               | 32,066             |
| <b>Panel B Neighborhood Characteristics</b> |                      |                      |                      |                    |
| Average Annual Income of Residents          | 73.0K<br>(58.0K)     | 80.7K<br>(73.7K)     | 72.5K<br>(55.2K)     | 8.2K<br>(.7K)      |
| Average Annual Income of Workers            | 41.6K<br>(16.7K)     | 50.1K<br>(26.4K)     | 40.5K<br>(15.0K)     | 9.6K<br>(.2K)      |
| <i>Number of ZIP Codes Per Entry</i>        | 114                  | 10                   | 60                   | 70                 |
| House Price Change Relative to $\tau=-1$    | -0.062<br>(0.290)    | -0.060<br>(0.307)    | -0.061<br>(0.272)    | 0.001<br>(0.002)   |
| <i>Number of Census Tracts Per Entry</i>    | 490                  | 39                   | 244                  | 283                |

Notes: Panels A and B report the summary statistics of our sample of incumbent residents and Zip Codes/Census Tracts, respectively, 1–4 years prior to the firm entry announcement. The sample of individuals consists of all the individuals who live within 30 minutes of driving time from a firm entry location in our sample as of Dec. 31st of the year prior to the announcement year. The sample of neighborhoods consists of all Zip Codes/Census Tracts within 30 minutes of driving time from each entry location. The “All”, “Inner Ring” and “Outer Ring” columns report the mean and standard deviation (in parentheses) of each outcome variable for all the individuals or neighborhoods within 0–30 mins (entry zone), 0–10 mins (inner ring), and 20–30 mins (outer ring) of 1990 level driving time from a firm entry, respectively. The “Difference” column reports the coefficient and standard error (in parentheses) of a regression of each outcome variable on the “Inner Ring” dummy. Number of residents (Zip Codes/Census Tracts) per entry reports the average number of incumbent residents (Zip Codes/Census Tracts) per firm entry in each group. Average incomes of residents and workers are reported at the ZIP Code level in 2010 dollars. Home value change, defined as the housing price index in each year minus the housing price index in the year before the entry announcement, is reported at the Census Tract level. Appendix Section C.3 documents how we construct this index.

Table 3: Heterogeneity of Treatment Effects on Neighborhood Outcomes across Entries

|                        | House Price Index<br>(1) | Log Median Rent<br>(2) | Workplace Income<br>(3) | Residential Income<br>(4) |
|------------------------|--------------------------|------------------------|-------------------------|---------------------------|
| Land Supply Elasticity | -0.0231**<br>(0.009)     | -0.00763<br>(0.006)    | -0.0109*<br>(0.007)     | -0.0150**<br>(0.006)      |
| High Skill Share       | 0.0724<br>(0.127)        | -0.0749<br>(0.073)     | 0.183*<br>(0.111)       | 0.0239<br>(0.095)         |
| Commuting Openness     | -0.0324<br>(0.062)       | 0.0157<br>(0.032)      | -0.00989<br>(0.051)     | -0.0388<br>(0.046)        |
| Pharmaceutical         | 0.00442<br>(0.030)       | -0.0109<br>(0.020)     | -0.0129<br>(0.028)      | 0.0185<br>(0.021)         |
| Energy and Equipment   | 0.0172<br>(0.030)        | -0.0314*<br>(0.018)    | 0.00119<br>(0.022)      | -0.00407<br>(0.017)       |
| Technology: Software   | 0.0484<br>(0.036)        | -0.0231<br>(0.021)     | -0.00737<br>(0.023)     | 0.0188<br>(0.021)         |
| Finance and Insurance  | 0.0461<br>(0.034)        | 0.0233<br>(0.021)      | -0.0208<br>(0.026)      | 0.0398*<br>(0.021)        |
| Other Service          | 0.0132<br>(0.042)        | -0.0118<br>(0.023)     | -0.0126<br>(0.031)      | -0.0102<br>(0.027)        |
| Technology: Hardware   | -<br>-                   | -<br>-                 | -<br>-                  | -<br>-                    |
| Constant               | 0.0287<br>(0.083)        | 0.0397<br>(0.030)      | -0.00993<br>(0.046)     | 0.0135<br>(0.039)         |
| No. Observations       | 289                      | 339                    | 264                     | 282                       |
| $R^2$                  | 0.0560                   | 0.0831                 | 0.0250                  | 0.0525                    |

Notes: We regress the estimated treatment effects on neighborhood outcomes from estimating equation (1) for each individual firm entry on characteristics of the entry zone and the firm. The dependent variables are estimated treatment effects for each individual entry on neighborhood outcomes for neighborhoods in the inner ring within 10 mins of driving time from the firm entry, relative to those in the outer rings 20–30 mins away from the entry, 6–10 years after the announcement. In column (1), the dependent variable is the estimated treatment effect on the house price index at the Census Tract level for each entry. In column (2), the dependent variable is the estimated treatment effect on log median rent at Census Tract level. In column (3), the dependent variable is the estimated treatment effect on log average income in a workplace ZIP Code. In column (4), the dependent variable is the estimated treatment effect on log average income in a residential ZIP Code. We drop firm entry cases where there are too few neighborhoods in the inner ring (fewer than 5 Census Tracts or 3 ZIP Codes) to credibly identify the treatment effects. The distributions of the estimated case-by-case treatment effects are plotted in Figure B10. The land supply elasticity is taken from [Saiz \(2010\)](#) for the closest CBSA to the entry location whenever it is available. This measure takes into account of both land availability due to geographic constraints and strictness of land-use regulations. The high skill share is the share of high-skilled workers who live within each entry zone prior to the entry. Following [Monte et al. \(2018\)](#), the commuting openness is defined as the share of residents within each entry zone who work outside each entry zone prior to the entry. The 6 industry categories of firm entry are defined in table A7. The omitted group of the industry categories is hardware technologies. The dependent variables are winsorized at the 1.0 and 99.5 percent level. Robust standard errors are presented in parentheses. \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ .

Table 4: GMM Estimates of Model Parameters

| <i>Estimated parameters</i>                               | High-skilled             | Low-skilled              |
|-----------------------------------------------------------|--------------------------|--------------------------|
| Preference for wage $\beta^s = \frac{1}{\sigma^s}$        | 1.335***<br>(0.372)      | 1.320***<br>(0.355)      |
| Preference for rent $-\beta^s\theta^R$                    | -0.828***<br>(0.230)     | -0.818***<br>(0.220)     |
| Preference for commute<br>cost $-\beta^s\lambda^s$        | -0.0216***<br>(0.000199) | -0.0212***<br>(0.000199) |
| Birthplace disamenity<br>born inside $-\sigma^s b_{in}$   | -5.063**<br>(2.51170)    | -5.121**<br>(2.40135)    |
| Birthplace disamenity<br>born outside $-\sigma^s b_{out}$ | -7.537**<br>(3.73892)    | -7.623**<br>(3.57466)    |
| <i>Calibrated parameters</i>                              |                          |                          |
| Housing expenditure share                                 | 0.62                     | 0.62                     |
| FE: Distance Bins $\times$ Entry Zone $\times$ Work ZIP   | Yes                      | Yes                      |

Notes: This table reports the estimated parameters for worker's preference for a sample of 129 firm entries during 2003–2010. Workers' preferences for wage and rent represent worker's demand elasticity with respect to the given neighborhood characteristic. Preference for commute cost represents worker's semi-elasticity of demand for a pair of residential and workplace neighborhoods with respect to one minute increase in the commute time. Birthplace disamenity estimates represent worker's semi-elasticity of demand for a neighborhood to whether that neighborhood is worker's initial location five years prior to the firm entry. In the second step of the estimation procedure, we add fixed effects for distance bins to nearby ZIP Codes with high predicted probability of attracting a high-skilled firm entry, interacted with the entry zone ID and workplace ZIP Code. Let  $\mathbf{Z}_{i,k} = [Z_{i,k}^1, Z_{i,k}^2]$  denote the vector of the closest distance from treated neighborhood  $i$  associated with firm entry  $k$  to ZIP Codes within 50 miles of the entry site of  $k$  in the quintiles,  $Q_1$  and  $Q_2$ , with the highest propensity scores, respectively. We then split neighborhoods into bins based on their closest distances to the top two propensity score quintiles. We split  $Z_{i,k}^1$  and  $Z_{i,k}^2$  into bins defined by 0–5, 5–10, and 10+ miles.

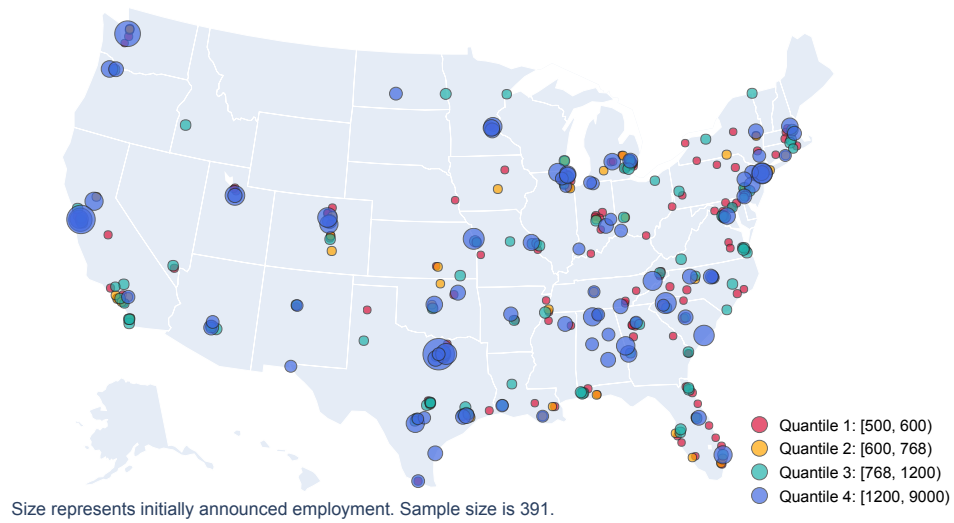
Table 5: Annual Welfare Change Due to Firm Entry: 5 Years After Entry

| <b>Panel A: No sorting in t=2</b>     |              |        |             |        |
|---------------------------------------|--------------|--------|-------------|--------|
|                                       | High-skilled |        | Low-skilled |        |
|                                       | Homeowner    | Renter | Homeowner   | Renter |
| <b><i>Total</i></b>                   | 109.6        | -32.0  | 20.6        | -55.3  |
| <b><i>Total (% Wage in t = 1)</i></b> | 0.21%        | -0.06% | 0.07%       | -0.20% |
| Wage                                  | 109.3        | 124.3  | 20.4        | 23.3   |
| Housing Dividend                      | 137.8        | 0      | 68.8        | 0      |
| Rent                                  | -137.4       | -156.4 | -68.7       | -78.5  |
| Others                                | 0            | 0      | 0           | 0      |
| <b>Panel B: Sorting in t=2</b>        |              |        |             |        |
| <b><i>Total</i></b>                   | 110.5        | -32.0  | 20.7        | -55.2  |
| <b><i>Total (% Wage in t = 1)</i></b> | 0.21%        | -0.06% | 0.07%       | -0.20% |
| Wage                                  | 262.7        | 56.9   | -18.3       | -103.2 |
| Housing Dividend                      | 137.8        | 0      | 68.8        | 0      |
| Rent                                  | -217.3       | -121.4 | -47.4       | -8.9   |
| Amenity                               | -50.4        | -5.6   | -24.0       | -6.2   |
| Others                                | -22.3        | 38.0   | 41.6        | 63.1   |

Notes: First row reports the annual average change in a worker's expected utility in real 2010 dollars across 129 firm entries, five years after the firm entry announcement. The second reports the annual welfare change in terms of percent change in workers' average wages in period 1. These welfare numbers can be interpreted as the welfare change for a representative high-skilled firm entry in our sample. In Panel A, we group the welfare changes due to commuting, local amenities, birthplace amenities and match-specific amenities into a single category of "Others". In Panel B, we group the welfare changes due to local amenities and match-specific amenities into a single category of "Amenity", and group commuting and birthplace disamenities into a single category of "Others".

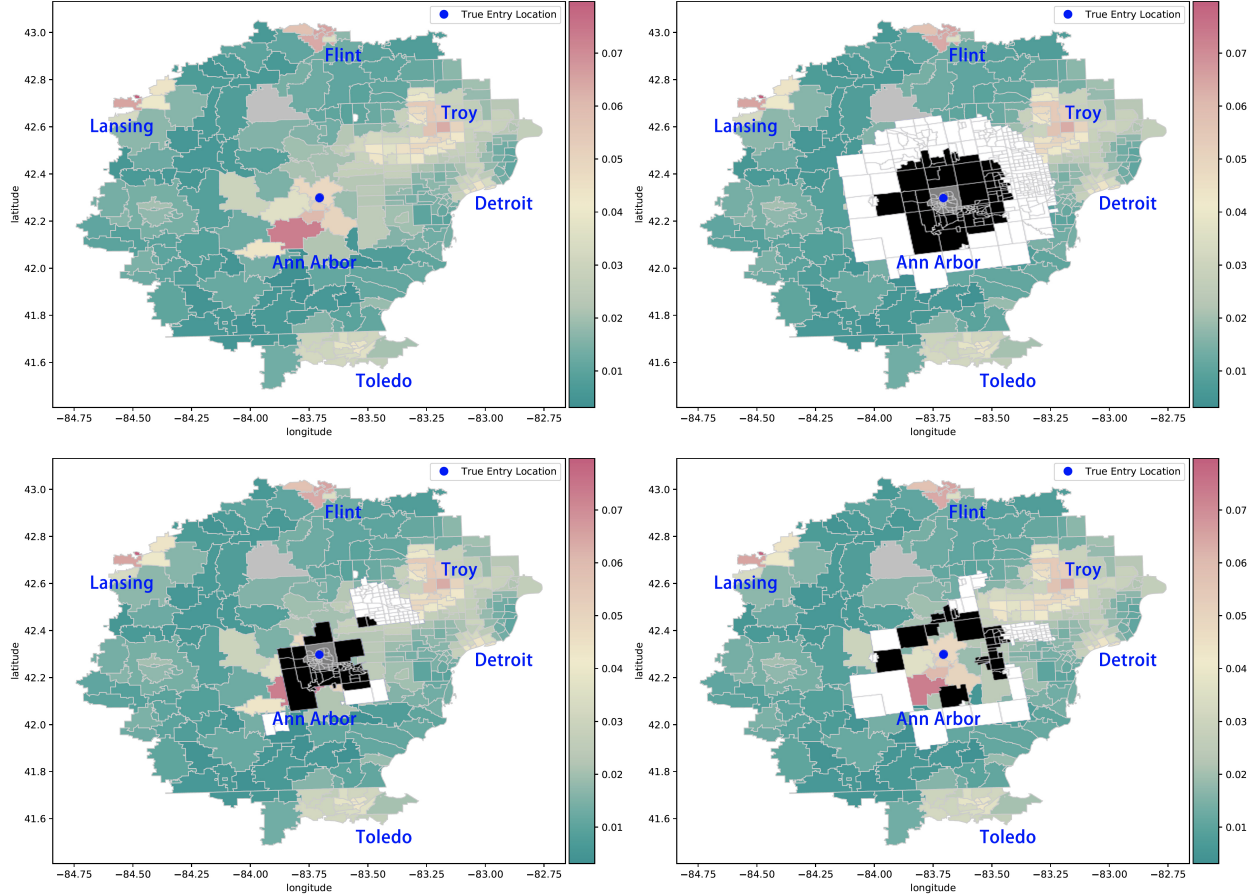


Figure 1: Sample of 391 High-Skilled Firm Entries



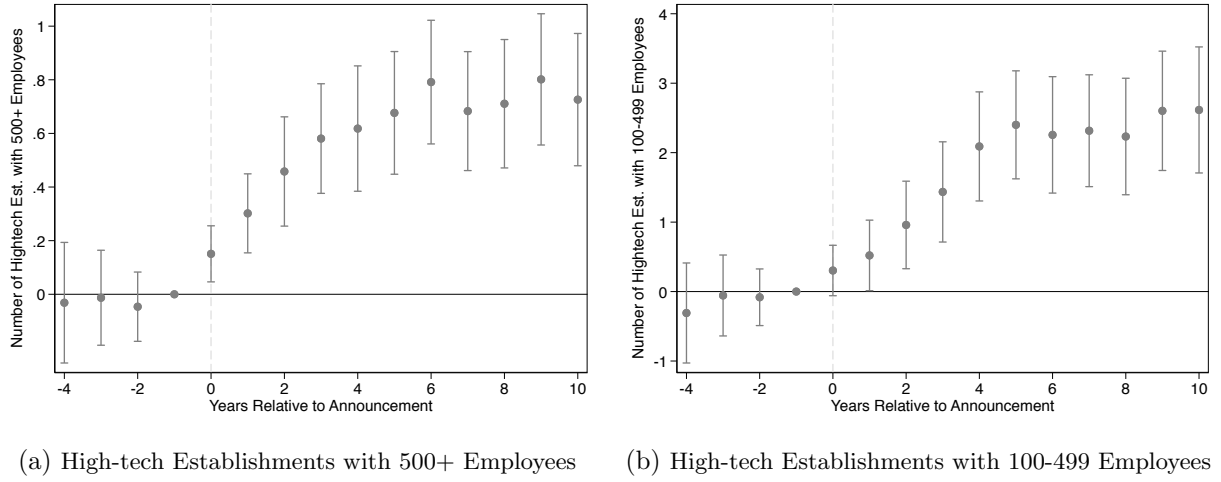
Notes: Each circle represents a firm entry. Entries that are located in the same city are at least 5 years apart. The size of the circles are proportional to the target employment, which is the number of jobs promised at the time the firm entry was announced. Firms that were announced but never opened are not part of this sample of 391 firms.

Figure 2: Map of Predicted Propensity Score of Firm Entry and Illustration of Identification



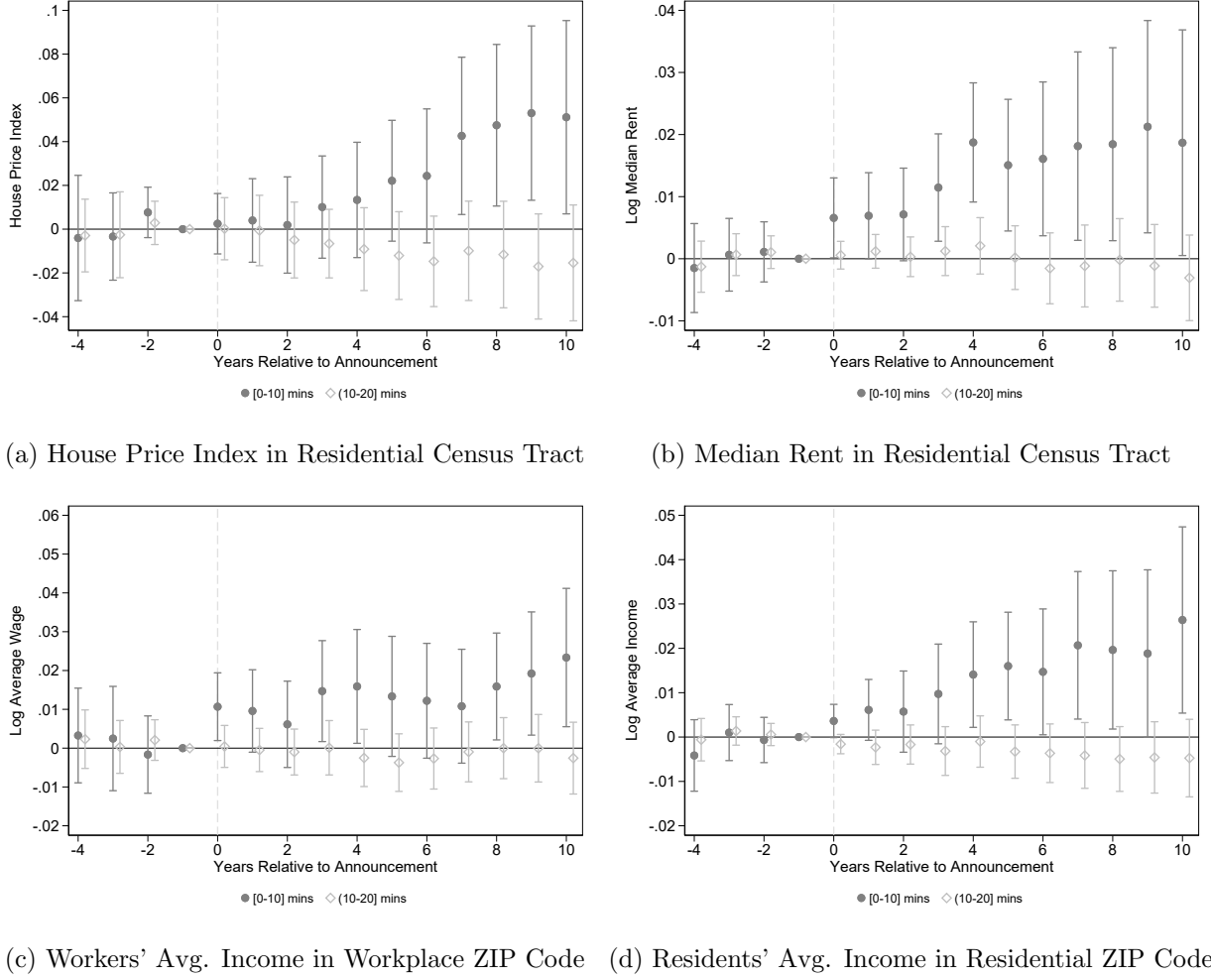
Notes: Panel (a) shows the predicted propensity score of high-skilled firm entry for ZIP Codes within 50 miles of the entry location of Pfizer in Ann Arbor, Michigan. Pfizer makes pharmaceutical products. The blue dot marks the actual location of Pfizer's entry. ZIP Codes with warmer colors have higher predicted probabilities of receiving a high-skilled firm entry. Areas marked by gray indicate there is not enough data to estimate the propensity score. Panel (b) shows the Census Tracts within 10, 10–20, and 20–30 minutes of driving time from the entry location. They are represented by the inner ring in gray, the middle ring in black and the outer ring in white. Panel (c) shows the Census Tracts in the inner, middle and outer ring whose closest distance (measured from Centroid to Centroid) to a ZIP Code in  $Q_1$ , the top quintile of ZIP Codes within 50 miles of the actual entry site with highest propensity scores is  $\leq 5$  miles (i.e., a potential entry site). These Census Tracts are similar in their proximity to nearby potential entry sites, hence similar in demographic and workplace characteristics that matter for a high-skilled firm's entry location. Panel (d) shows the Census Tracts in the middle and outer ring whose closest distance to a ZIP Code among the top-quintile potential entry sites is 5–10 miles. Panels (c) and (d) illustrate the example groups of neighborhoods defined by bins  $h_{i,k}^1$  in equation (1). In practice, equation (1) also controls for bins  $h_{i,k}^2$ , which are based on a neighborhood's closest distance to a ZIP Code in  $Q_2$ , the second top quintile of ZIP Codes within 50 miles of the actual entry site with highest propensity scores.

Figure 3: Validation of Propensity Score Model



Notes: This figure plots the coefficient of the event study to validate our propensity score model in equation (30). The outcome variable is the number of high-tech establishments with 500+ employees at ZIP codes level in Panel (a) and the number of high-tech establishments with 100-499 employees at ZIP codes level in Panel (b). The data source is ZIP Code Business Patterns. The treatment group are ZIP Codes that actually received a firm entry from our sample of 391 firm entries. The control group are ZIP Codes that did not receive an entry in our sample, but are among the top quintile of ZIP Codes within 50 miles of the actual entry location that have the highest predicted probability of receiving a high-skilled firm entry. 95% confidence intervals are provided, where standard errors are clustered at the entry zone level.

Figure 4: Event Study: The Effects of High-skilled Firm Entries on Neighborhood Outcomes

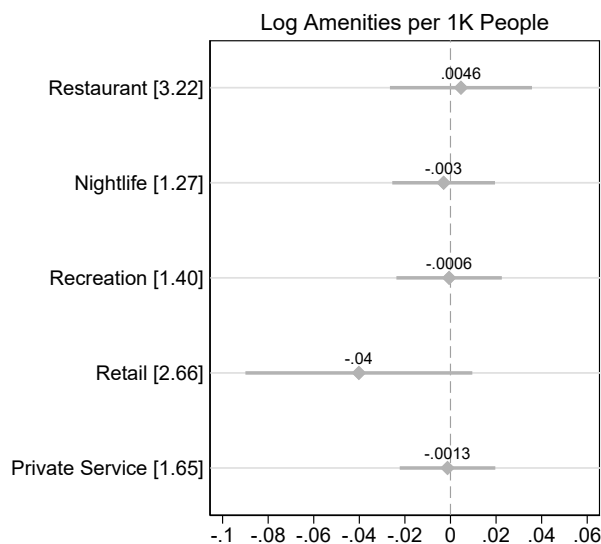


Notes: This figure plots the event study coefficients of treatment effects on several neighborhood outcomes for neighborhoods in the inner ring within 10 mins from the firm entry, relative to those in the outer rings 20–30 mins away from the entry site, before/after the entry announcement. The house price index at the Census Tract level is constructed using CoreLogic transactions records. Appendix C.3 documents how we construct this index. Log median rent at Census Tract level is from Censuses/ACS. Log average income in a workplace ZIP Code is computed using the average ZIP Code level Payroll data from ZIP Code Business Patterns. Log average income in a residential ZIP Code is computed using the average ZIP Code level adjusted gross income (AGI) from the IRS Statistics of Income Tax Stats. 95% confidence intervals are provided, where standard errors are clustered at the entry zone level. Figure B8 provides robustness check where we additionally control for a neighborhood's proximity to ZIP Codes with high prior density of high-tech employment.

Figure 5: The Effects of High-skilled Firm Entries on Neighborhood Outcomes 6–10 Years Later



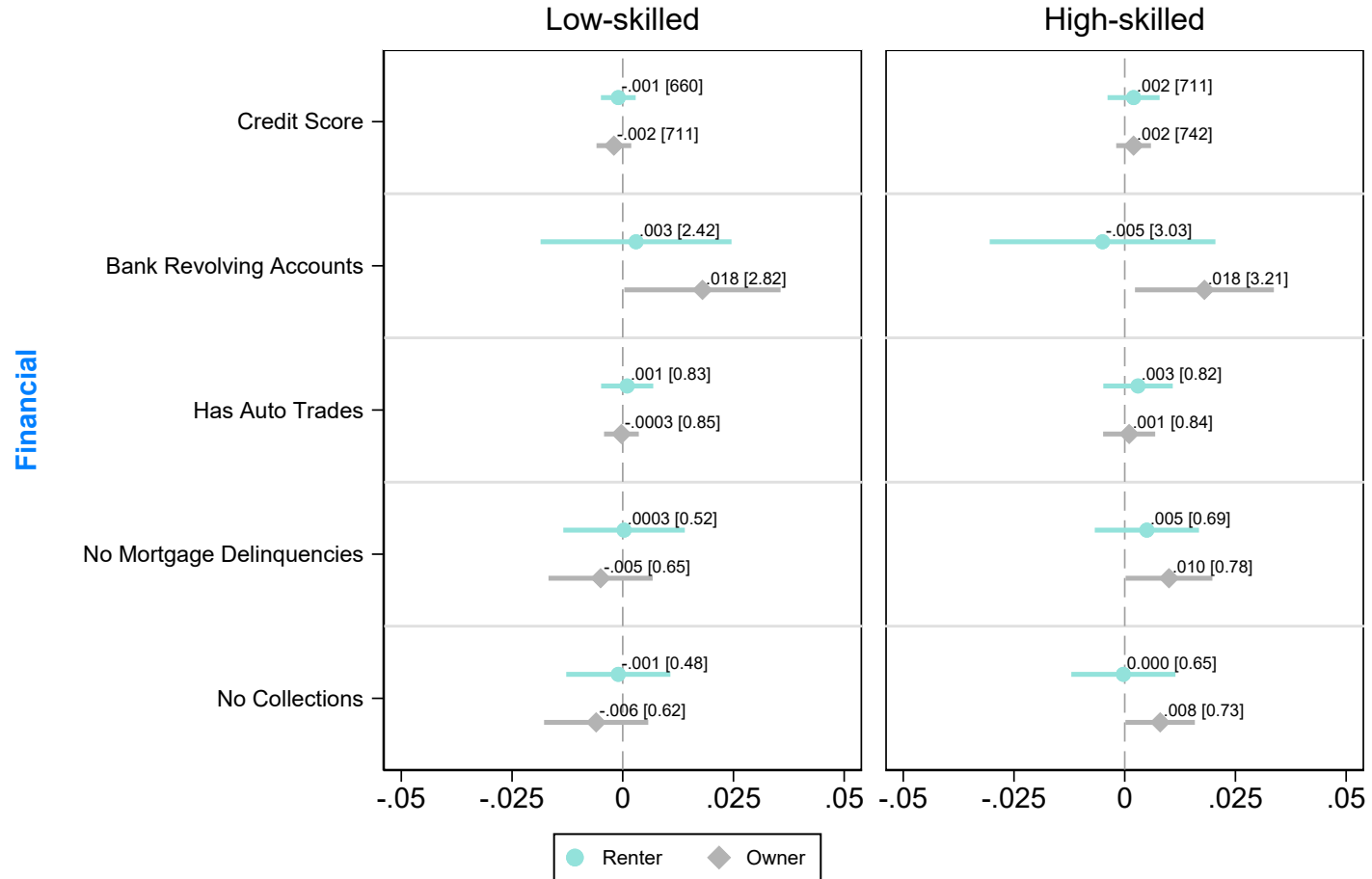
(a) House Prices and Income



(b) Amenities

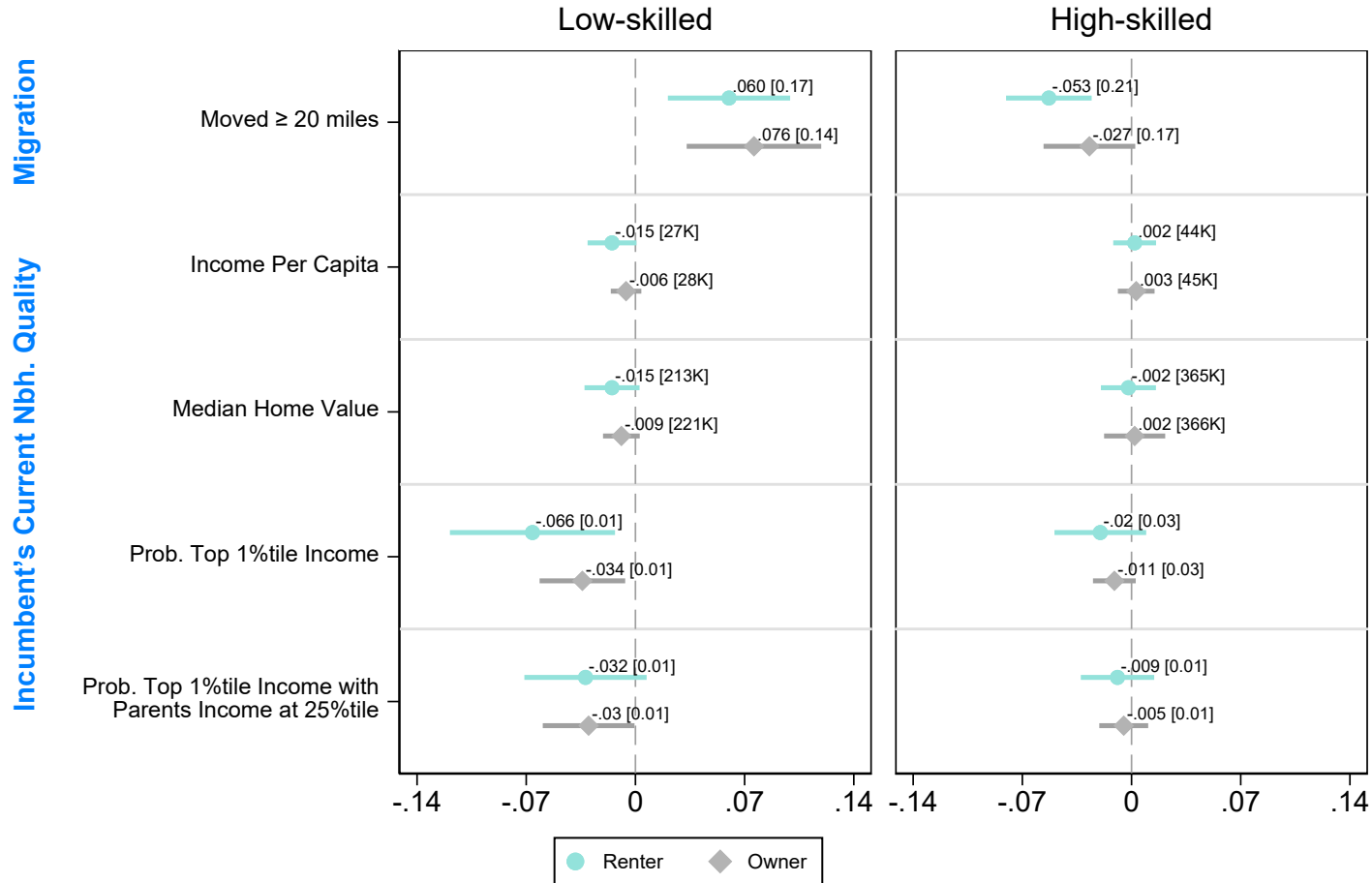
Notes: This figure plots the average treatment effects on several neighborhood outcomes for neighborhoods in the inner ring within 10 mins from the firm entry, relative to those in the outer rings 20–30 mins away from the entry site, 6–10 years after the entry announcement. In Panel (a), the house price index at Census Tract level is constructed using CoreLogic transactions data. Appendix C.3 documents how we construct this index. Log median rent at Census Tract level is from Censuses/ACS. Log average income in a workplace ZIP Code is computed using the average ZIP Code level Payroll data from ZIP Code Business Patterns. Log average income in a residential ZIP Code is computed using the average ZIP Code level adjusted gross income (AGI) from the IRS Statistics of Income Tax Stats. Panel (b) plots the average treatment effects on ZIP Code level amenities 6–10 years later. Table A15 documents how we construct measures of these amenities. We present the mean of the dependent variables in the control group 6–10 years after the announcement in brackets. 95% confidence intervals are provided, where standard errors are clustered at the entry zone level. Figure B9 provides robustness check where we additionally control for a neighborhood's proximity to ZIP Codes with high prior density of high-tech employment.

Figure 6: Financial Outcomes



Notes: Treatment effects 5 to 8 years after firm entry announcement are in percent terms of the 0–10 minute ring around the entry site relative to the mean of the control group, which is the 20–30 minute ring around the entry site. Treatment effects are presented for four groups: low-skilled renters, low-skilled owners, high-skilled renters, and high-skilled owners. The lines to either side of the green circle and gray diamond for point estimates represent 95% confidence intervals. Control means of each group 5 to 8 years after firm entry announcement are in brackets next to the point estimates. The first outcome variable is the credit score. The second outcome variable is the number of bank revolving accounts, which proxy for non-durable consumption. The third outcome variable is whether an individual has ever had an auto trade, which proxy for durable consumption. The fourth outcome variable is whether an individual has never had a mortgage delinquency. The fifth outcome variable is whether an individual has never had a debt collection.

Figure 7: Share Moved and Quality of Incumbent's Current Neighborhood Outcomes



Notes: Treatment effects 5 to 8 years after firm entry announcement are in percent terms of the 0–10 minute ring around the entry site relative to the mean of the control group, which is the 20–30 minute ring around the entry site. Treatment effects are presented for four groups: low-skilled renters, low-skilled owners, high-skilled renters, and high-skilled owners. The lines to either side of the green circle and gray diamond for point estimates represent 95% confidence intervals. Control means of each group 5 to 8 years after firm entry announcement are in brackets next to the point estimates. The first outcome variable is the share of incumbent residents who moved more than 20 miles away from their address when the firm entry is announced, which proxies for moving out of the city. The second and third outcome variables are the contemporaneous income per capita and the contemporaneous median home value of the Census Tract where individuals reside. The last two outcome variables are the upward mobility measures from [Chetty et al. \(2018\)](#) and based on the sample of children in the 1978–1983 birth cohorts who are U.S. citizen or authorized immigrants. Children's individual incomes are measured in 2014 and 2015, when they are between the ages of 31 and 37. The fourth outcome variable is the mean probability of reaching the top 1% of the national individual income distribution which is not conditioning on the parental income. The fifth outcome variable is the probability of reaching the top 1% of the national individual income distribution conditioning on the parental income at 25 percentile in the national household income distribution.

# Appendices

|                                                                           |            |
|---------------------------------------------------------------------------|------------|
| <b>A Additional Tables</b>                                                | <b>50</b>  |
| <b>B Additional Figures</b>                                               | <b>65</b>  |
| <b>C Data Construction</b>                                                | <b>94</b>  |
| C.1 Homeownership Imputation . . . . .                                    | 94         |
| C.2 Education/Skill Imputation . . . . .                                  | 96         |
| C.3 Census Tract Level House Price Indices . . . . .                      | 98         |
| C.4 Driving Time . . . . .                                                | 99         |
| C.5 Data Implementation for Estimation . . . . .                          | 100        |
| C.6 10 Cases of Firm Entry . . . . .                                      | 105        |
| <b>D Propensity Score Model</b>                                           | <b>107</b> |
| <b>E Model</b>                                                            | <b>109</b> |
| E.1 Worker’s Choice for Residential and Workplace Neighborhoods . . . . . | 109        |
| E.2 Labor Supply . . . . .                                                | 110        |
| E.3 Labor Demand . . . . .                                                | 110        |
| E.4 Housing Supply . . . . .                                              | 111        |
| E.5 Housing Demand . . . . .                                              | 112        |
| E.6 Equilibrium . . . . .                                                 | 112        |
| E.7 Welfare Incidence of a High-Skilled Firm Entry . . . . .              | 112        |
| E.8 Simulation for a High-Skilled Firm Entry . . . . .                    | 113        |
| E.9 Estimation Procedure . . . . .                                        | 114        |
| E.10 Change in Expected Welfare for Movers vs. Stayers . . . . .          | 119        |

## A Additional Tables



Table A1: Infutor-CoreLogic: Address Merge Rates

| State | Infutor     | CoreLogic   | Matches    | Match rate |
|-------|-------------|-------------|------------|------------|
| AK    | 549,804     | 577,030     | 178,778    | 0.33       |
| AL    | 3,686,368   | 3,381,063   | 1,640,014  | 0.44       |
| AR    | 2,264,192   | 2,774,138   | 1,117,671  | 0.49       |
| AZ    | 4,084,237   | 5,895,776   | 2,140,958  | 0.52       |
| CA    | 21,752,607  | 22,692,583  | 10,271,136 | 0.47       |
| CO    | 3,637,375   | 4,882,710   | 1,831,535  | 0.5        |
| CT    | 2,254,395   | 501,270     | 219,599    | 0.1        |
| DC    | 519,947     | 403,199     | 177,154    | 0.34       |
| DE    | 673,260     | 814,861     | 341,318    | 0.51       |
| FL    | 13,834,611  | 20,195,793  | 7,597,739  | 0.55       |
| GA    | 6,545,877   | 7,749,462   | 3,314,527  | 0.51       |
| HI    | 991,644     | 1,189,447   | 389,401    | 0.39       |
| IA    | 2,114,242   | 2,559,005   | 1,188,525  | 0.56       |
| ID    | 1,178,220   | 1,455,100   | 554,947    | 0.47       |
| IL    | 8,479,168   | 9,809,275   | 4,273,576  | 0.5        |
| IN    | 4,620,489   | 5,518,307   | 2,369,448  | 0.51       |
| KS    | 1,977,176   | 2,217,902   | 1,040,895  | 0.53       |
| KY    | 3,390,553   | 3,116,883   | 1,580,044  | 0.47       |
| LA    | 3,326,382   | 3,391,884   | 1,641,317  | 0.49       |
| MA    | 4,129,508   | 960,665     | 380,885    | 0.09       |
| MD    | 3,587,829   | 4,640,438   | 1,890,075  | 0.53       |
| ME    | 1,250,021   | 139,026     | 30,400     | 0.02       |
| MI    | 7,000,246   | 5,635,832   | 3,728,442  | 0.53       |
| MN    | 3,447,233   | 4,290,624   | 1,895,469  | 0.55       |
| MO    | 4,307,747   | 4,990,370   | 2,150,686  | 0.5        |
| MS    | 2,434,565   | 1,901,738   | 969,713    | 0.4        |
| MT    | 801,754     | 1,084,111   | 350,949    | 0.44       |
| NC    | 7,072,070   | 9,315,465   | 3,586,847  | 0.51       |
| ND    | 522,541     | 483,394     | 200,736    | 0.38       |
| NE    | 1,247,222   | 1,394,951   | 643,619    | 0.52       |
| NH    | 1,043,455   | 132,807     | 47,363     | 0.05       |
| NJ    | 5,509,166   | 714,524     | 233,778    | 0.04       |
| NM    | 1,499,615   | 1,471,116   | 625,330    | 0.42       |
| NV    | 1,707,378   | 2,348,467   | 907,666    | 0.53       |
| NY    | 12,511,117  | 11,627,570  | 4,742,301  | 0.38       |
| OH    | 8,292,963   | 10,086,281  | 4,228,038  | 0.51       |
| OK    | 2,910,524   | 3,106,146   | 1,253,660  | 0.43       |
| OR    | 2,705,486   | 3,220,908   | 1,358,130  | 0.5        |
| PA    | 8,807,052   | 10,301,095  | 4,532,221  | 0.51       |
| RI    | 685,590     | 313,251     | 127,535    | 0.19       |
| SC    | 3,350,214   | 4,473,437   | 1,835,782  | 0.55       |
| SD    | 584,486     | 448,048     | 262,306    | 0.45       |
| TN    | 4,439,823   | 5,608,419   | 2,379,854  | 0.54       |
| TX    | 15,942,543  | 19,570,551  | 7,930,927  | 0.5        |
| UT    | 1,706,935   | 2,481,290   | 812,449    | 0.48       |
| VA    | 5,216,557   | 6,111,974   | 2,758,078  | 0.53       |
| VT    | 581,643     | 59,039      | 3,748      | 0.01       |
| WA    | 4,987,250   | 5,584,186   | 2,331,650  | 0.47       |
| WI    | 4,433,949   | 5,279,060   | 2,024,908  | 0.46       |
| WV    | 1,560,793   | 1,582,449   | 468,113    | 0.3        |
| WY    | 439,529     | 512,887     | 195,221    | 0.44       |
| Total | 210,597,351 | 228,995,807 | 96,755,461 | 0.46       |

Notes: This table shows merge rates between unique addresses in Infutor and Core Logic deeds and tax records by state. The Infutor and CoreLogic columns shows the number of distinct addresses in Infutor and CoreLogic respectively (after cleaning the addresses as described in Appendix C.1.1). The match rate refers to the fraction of Infutor addresses matched to CoreLogic.

Table A2: Prediction of Skill/Education using ACS

|                                           | (1)                      | (2)                      | (3)                  |
|-------------------------------------------|--------------------------|--------------------------|----------------------|
| <b><i>Housing Characteristics</i></b>     |                          |                          |                      |
| Log of House Value                        | 0.837***<br>(0.003)      | 0.828***<br>(0.008)      | 0.668***<br>(0.236)  |
| Log of Rent                               | 1.304***<br>(0.009)      | 1.412***<br>(0.017)      | 2.560***<br>(0.829)  |
| No. Bedrooms                              | 0.117***<br>(0.006)      | 0.103***<br>(0.027)      | -0.0434<br>(0.349)   |
| No. Rooms                                 | -0.0324***<br>(0.003)    | 0.0324**<br>(0.014)      | 0.107<br>(0.149)     |
| Bedroom/Room                              | -2.081***<br>(0.043)     | -1.601***<br>(0.171)     | -1.591<br>(2.590)    |
| Building Age                              | -0.00300***<br>(0.001)   | 0.0191***<br>(0.003)     | 0.0229<br>(0.047)    |
| Building Age <sup>2</sup>                 | -0.000185***<br>(0.000)  | -0.000638***<br>(0.000)  | -0.00125<br>(0.001)  |
| Building Age <sup>3</sup>                 | 0.00000270***<br>(0.000) | 0.00000445***<br>(0.000) | 0.0000120<br>(0.000) |
| Not Single-family Nor Condo               | -0.0150**<br>(0.006)     | -0.0771<br>(0.455)       | -0.982<br>(0.747)    |
| Rented                                    | 0.897***<br>(0.074)      | 2.325***<br>(0.423)      | -7.374<br>(6.425)    |
| <b><i>Demographic Characteristics</i></b> |                          |                          |                      |
| Female                                    | 0.0872***<br>(0.004)     | 2.968***<br>(0.266)      | 2.271***<br>(0.391)  |
| Immigrant                                 | 0.224***<br>(0.008)      | 2.787***<br>(0.581)      | 2.913***<br>(0.812)  |
| Marriage                                  |                          |                          |                      |
| Separate, Divorced, Widowed               | -0.231***<br>(0.006)     | -2.852***<br>(0.495)     | -2.498***<br>(0.645) |
| Single                                    | 0.120***<br>(0.006)      | 2.383***<br>(0.355)      | 1.470***<br>(0.547)  |
| Race                                      |                          |                          |                      |
| Non-Hispanic Black                        | -0.795***<br>(0.012)     | 2.482***<br>(0.771)      | 2.721**<br>(1.248)   |
| Hispanic                                  | -1.457***<br>(0.011)     | 3.583***<br>(0.720)      | 3.889***<br>(1.161)  |
| Non-Hispanic White                        | -0.394***<br>(0.010)     | 4.630***<br>(0.629)      | 4.549***<br>(1.063)  |
| Age                                       | 0.191***<br>(0.009)      | 0.770***<br>(0.046)      | 0.428<br>(0.617)     |
| Age <sup>2</sup>                          | -0.00531***<br>(0.000)   | -0.0190***<br>(0.001)    | -0.0139<br>(0.014)   |
| Age <sup>3</sup>                          | 0.0000427***<br>(0.000)  | 0.000145***<br>(0.000)   | 0.000123<br>(0.000)  |
| Observations                              | 1,439,928                | 1,439,928                | 1,439,928            |
| Pseudo R-Squared                          | 0.130                    | 0.153                    | 0.159                |
| Accuracy (Cutoff 0.5)                     | 72.42                    | 73.59                    | 73.78                |
| State FE                                  | No                       | Yes                      | Yes                  |
| Interactions with States FE               | No                       | No                       | Yes                  |

Notes: Sample consists of adults aged between 25 and 65 from the 1% sample of 2008-2012 5-year ACS. A person is defined to be high-skilled if he holds a bachelor degree or above. We stratify the ACS sample by county, and divide 90% of the ACS sample as training data and 10% as test data. Columns 1 reports the baseline logit model using our training data from ACS; Columns 2 adds state fixed effects and interaction terms between independent variables; Columns 3 further adds the interaction terms between independent variables and state dummies. Robust standard errors are reported in parentheses. For property type, the omitted category is single family or condominium; For Gender, the omitted category is Male; for marriage, the omitted category is married; For race, the omitted category is non-Hispanic Asian; For immigration status, the omitted category is non-immigrant or native. In our Infutor sample, we define an immigrant as a person who immigrated to US at age 20 or later. The missing value dummies, state fixed effects and interaction terms are omitted in the table. \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ .

Table A3: Census Block Group Skill Distribution by Worker’s Predicted Skill Level Among Sample of Individuals

|                                  |              | US Overall |          | Avg. High-Skilled Share    |
|----------------------------------|--------------|------------|----------|----------------------------|
|                                  | Sample Share | 2010 ACS   | 2015 ACS | in 2015 Census Block Group |
|                                  | (1)          | (2)        | (3)      | (4)                        |
| <i>Predicted Skill/Education</i> |              |            |          |                            |
| High                             | 29.55        | 29.87      | 32.01    | 55.76                      |
| Low                              | 70.45        | 70.13      | 67.99    | 28.11                      |

Notes: Column 1 reports the share of predicted high- and low-skilled workers in our sample. Columns 2 and 3 report the nation-wide shares of high- and low-skilled workers between 25 and 65 years old according to the 2010 and 2015 1-year ACS. Column 4 reports the average share of high-skilled residents in the Census block group where an individual resides in 2015. The sample consists of individuals in the Infutor-CoreLogic linked sample between 20 and 65 years old living within 30 minutes of the firm entry location as of December 31 of the year before firm entry announcement. Housing characteristics are from the individual’s address in December 2010.

Table A4: Match of Infutor Sample to TU Credit Reports

|               | (1)<br>TU Match       |
|---------------|-----------------------|
| 10–20 min     | -0.00025<br>(0.00025) |
| 20–30 min     | -0.00011<br>(0.00033) |
| Entry Zone FE | Yes                   |
| Observations  | 120,657,004           |

Notes: This tables shows the regression coefficients when we regress the indicator of whether an Infutor panelist in our main sample is matched to TU credit reports on a set of dummies of initial locations or treatment group status, i.e., whether they initially live within 10 mins of driving time, 10–20 mins, or 20–30 mins away from a firm entry location. The omitted category is within 10 mins. The regression includes entry zone fixed effects and standard errors are clustered at the entry zone level. \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ .

Table A5: Estimated Parameters of Propensity Score Model

| Variable                                                                           | Coefficient |
|------------------------------------------------------------------------------------|-------------|
| Constant                                                                           | -8.77       |
| Hightech job density 0-5 mi                                                        | 1.10        |
| Hightech job density 5-10 mi                                                       | 0.27        |
| Hightech job density 20-30 mi                                                      | -0.86       |
| Job density in Chemicals & Allied Products 0-5 mi                                  | 0.16        |
| Job density in Computers, Software, Electronic Equipment 0-5 mi                    | 2.31        |
| Job density in Computers, Software, Electronic Equipment 5-10 mi                   | 2.34        |
| Job density in Mines, Constr, BldMt, Trans, Hotels, Bus Serv, Entertainment 0-5 mi | -0.42       |
| Job density in Food, Tobacco, Textiles, Apparel, Leather, Toys 10-20 mi            | -0.15       |
| Job density in Utilities 10-20 mi                                                  | -0.22       |
| Job density in Healthcare, Medical Equipment, Drugs 20-30 mi                       | -1.81       |
| Log average workplace wage 0-5 mi                                                  | 0.39        |
| Employment/population 0-5 mi $\times$ % population age 19- 0-5 mi                  | 3.36        |
| Employment/population 0-5 mi $\times$ % Unemployed 20-30 mi                        | 4.47        |
| Employment/population 0-5 mi $\times$ % Poor 10-20 mi                              | 0.38        |
| % Population age 20-44 10-20 mi                                                    | 0.01        |
| % Population age 65+ 0-5 mi                                                        | -0.19       |
| % Population age 65+ 5-10 mi                                                       | -3.52       |
| % Household heads moved into unit less than 10 years 0-5 mi                        | 3.88        |

Notes: This tables shows the estimated parameters of the propensity score model for the conditional probability for a ZIP Code to receive a high-skill firm entry given the observed characteristics of itself and its surrounding neighborhoods in 1990, prior to any firm entries in our sample.

Table A6: Effect on High-skilled Job Growth

| Control group<br>Dependent variable                                        | Top Quintile ZIPs                         |                                        | Top Decile ZIPs                           |                                        |
|----------------------------------------------------------------------------|-------------------------------------------|----------------------------------------|-------------------------------------------|----------------------------------------|
|                                                                            | # Hightech Est. with<br>100–499 Employees | # Hightech Est. with<br>500+ Employees | # Hightech Est. with<br>100–499 Employees | # Hightech Est. with<br>500+ Employees |
|                                                                            | (1)                                       | (2)                                    | (3)                                       | (4)                                    |
| <b>Panel A Baseline Results</b>                                            |                                           |                                        |                                           |                                        |
| $\tau = 0-5 \times \text{Treat}$                                           | 1.173***<br>(0.264)                       | 0.447***<br>(0.079)                    | 1.125***<br>(0.266)                       | 0.436***<br>(0.08)                     |
| $\tau = 6-10 \times \text{Treat}$                                          | 2.475***<br>(0.378)                       | 0.748***<br>(0.101)                    | 2.404***<br>(0.381)                       | 0.737***<br>(0.103)                    |
| Avg. Change in dep. var.<br>in control ZIPs $\tau = -1-10$                 | 0.43                                      | 0.11                                   | 0.61                                      | 0.13                                   |
| Metro Area $\times$ ZIP FE                                                 | Yes                                       | Yes                                    | Yes                                       | Yes                                    |
| Metro Area $\times$ Year FE                                                | Yes                                       | Yes                                    | Yes                                       | Yes                                    |
| Adjusted $R^2$                                                             | 0.91                                      | 0.80                                   | 0.91                                      | 0.81                                   |
| Observations                                                               | 214,858                                   | 214,858                                | 106,295                                   | 106,295                                |
| <b>Panel B Compare Actual &amp; Potential<br/>Sites Within Same County</b> |                                           |                                        |                                           |                                        |
| $\tau = 0-5 \times \text{Treat}$                                           | 0.995***<br>(0.318)                       | 0.483***<br>(0.095)                    | 0.922***<br>(0.348)                       | 0.425***<br>(0.104)                    |
| $\tau = 6-10 \times \text{Treat}$                                          | 2.489***<br>(0.481)                       | 0.757***<br>(0.127)                    | 2.396***<br>(0.531)                       | 0.752***<br>(0.146)                    |
| Avg. Change in dep. var.<br>in control ZIPs $\tau = -1-10$                 | 0.53                                      | 0.15                                   | 0.61                                      | 0.15                                   |
| Metro Area $\times$ ZIP FE                                                 | Yes                                       | Yes                                    | Yes                                       | Yes                                    |
| Metro Area $\times$ Year FE                                                | Yes                                       | Yes                                    | Yes                                       | Yes                                    |
| Adjusted $R^2$                                                             | 0.91                                      | 0.79                                   | 0.91                                      | 0.79                                   |
| Observations                                                               | 54,738                                    | 54,738                                 | 28,638                                    | 28,638                                 |
| <b>Panel C P-score Model<br/>with Decadal Changes</b>                      |                                           |                                        |                                           |                                        |
| $\tau = 0-5 \times \text{Treat}$                                           | 1.113***<br>(0.294)                       | 0.447***<br>(0.089)                    | 1.085***<br>(0.295)                       | 0.451***<br>(0.090)                    |
| $\tau = 6-10 \times \text{Treat}$                                          | 2.417***<br>(0.432)                       | 0.770***<br>(0.113)                    | 2.408***<br>(0.435)                       | 0.797***<br>(0.115)                    |
| Avg. Change in dep. var.<br>in control ZIPs $\tau = -1-10$                 | 0.43                                      | 0.10                                   | 0.47                                      | 0.08                                   |
| Metro Area $\times$ ZIP FE                                                 | Yes                                       | Yes                                    | Yes                                       | Yes                                    |
| Metro Area $\times$ Year FE                                                | Yes                                       | Yes                                    | Yes                                       | Yes                                    |
| Adjusted $R^2$                                                             | 0.91                                      | 0.80                                   | 0.91                                      | 0.82                                   |
| Observations                                                               | 187,844                                   | 187,844                                | 93,244                                    | 93,244                                 |

Notes: This table summarizes the results of the event study to validate our propensity score model using equation (30). It summarizes the event study coefficients in the post-entry periods. The full set of event study coefficients are presented in Figure 3 and Figure B6. Panel A shows the baseline results using our baseline p-score model reported in the paper. Panel B shows the results by comparing actual sites that received a firm entry with top potential sites within the same county rather than within the same metropolitan area as in the baseline model. This further controls for potential unobserved differences between treated and control sites that could differ across counties within the same metropolitan area, such as differences in tax incentives. Panel C shows the results after adding decadal changes of neighborhood demographic and workplace characteristics into the p-score model. This further controls for dynamic aspects of neighborhood changes that could affect firms' location decisions. For the firm entries that happened after 2000, we use 1990–2000 decadal changes of neighborhood characteristics. For entries happened during 1990s, we use 1980–1990 changes. The treatment group are ZIP Codes that actually received a firm entry from our sample. In columns (1) and (2), the control group are ZIP Codes that did not receive an entry in our sample, but are among the top quintile of ZIP Codes within 50 miles of the actual entry location that have the highest predicted probability of receiving a high-skilled firm entry. In columns (3) and (4), the control group are ZIP Codes from the top decile. In columns (1) and (3), the dependent variable is number of high-tech establishments of size 100–499 in a ZIP Code measured from the ZIP Code Business Patterns. In columns (2) and (4), the dependent variable is number of high-tech establishments of size 500+ measured from the ZIP Code Business Patterns. Standard errors are clustered at the metropolitan area (50-mile ring around the actual entry location) level. \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ .

Table A7: Definition of Industry Categories

| Industry Category     | Industries                                                       | Examples                                             |
|-----------------------|------------------------------------------------------------------|------------------------------------------------------|
| Pharmaceutical        | Biopharmaceutical, Pharmaceuticals, Health, Chemicals, Food      | Pfizer, Cephalon, Dow Chemical Company               |
| Energy and Equipment  | Energy, Utilities, Airplane, Automobile, Equipment Manufacturing | Exxon, GE, Boeing, Honeywell, ABB, Toyota, Honda, GM |
| Technology: Hardware  | Semiconductors, Computer, Devices                                | Intel, Texas Instruments, Dell, Motorola             |
| Technology: Software  | Telecommunication, Software, Media                               | Amazon, Google, Adobe, Qualcomm, Cisco, IBM          |
| Finance and Insurance | Bank, Insurance, Accounting, Law, Consulting                     | Fidelity, JP Morgan, Morgan Stanley                  |
| Other Service         | Travel, Logistics, Retail, Telemarketing                         | Best Buy, UPS, Interim, Sykes, Sabre, TeleTech       |

Notes: This table shows the definition of industry categories used in table 3 and example firm entries from each category.

Table A8: Model Calibration

|                             | Description                                     | Value       |
|-----------------------------|-------------------------------------------------|-------------|
| $\beta^H$                   | High-skilled preference for real wage           | 1.67        |
| $\beta^L$                   | Low-skilled preference for real wage            | 1.25        |
| $\lambda^H$                 | High-skilled commuting cost per mile            | 0.2         |
| $\lambda^L$                 | Low-skilled commuting cost per mile             | 0.1         |
| $a^H$                       | High-skilled amenity value                      | 0.2         |
| $a^L$                       | Low-skilled amenity value                       | 0           |
| $b_{in}$                    | Birthplace amenity if born inside               | 4           |
| $b_{out}$                   | Birthplace amenity if born outside              | 18          |
| $\alpha$                    | Labor share                                     | 0.66        |
| $\sigma = \frac{1}{1-\rho}$ | Elasticity of labor substitution                | 1.6         |
| $\kappa$                    | Interest rate                                   | 0.05        |
| $\theta^R$                  | Renter expenditure share of housing consumption | 0.62        |
| $\theta^O$                  | Owner expenditure share of housing consumption  | 0.62        |
| $k_i$                       | Inverse housing supply elasticity               | 0.8         |
| $X_{j,1}^H$                 | Initial high-skilled productivity               | $\exp(1)$   |
| $X_{j,1}^L$                 | Initial low-skilled productivity                | $\exp(0.8)$ |
| $\Delta \ln X_j^H$          | Change in high-skilled log productivity         | 0.2         |

Notes: This table shows the calibrated values of model parameters for the simulation of our full spatial equilibrium model.

Table A9: Candidate Explanatory Variables in Model of Commute Pattern by Skill

| Source     | Variables                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>OD</b>  | % jobs of workers age 29 or younger<br>% jobs for workers age 30 to 54<br>% jobs with earnings \$1250/month or less<br>% jobs with earnings \$1251/month to \$3333/month<br>% jobs in Goods Producing industry sectors<br>% jobs in Trade, Transportation, and Utilities industry sector                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| <b>RAC</b> | % jobs for workers age 29 or younger<br>% jobs for workers age 30 to 54<br>% jobs with earnings \$1250/month or less<br>% jobs with earnings \$1251/month to \$3333/month<br>% jobs in NAICS sector 11 (Agriculture, Forestry, Fishing and Hunting)<br>% jobs in NAICS sector 21 (Mining, Quarrying, and Oil and Gas Extraction)<br>% jobs in NAICS sector 22 (Utilities)<br>% jobs in NAICS sector 23 (Construction)<br>% jobs in NAICS sector 31-33 (Manufacturing)<br>% jobs in NAICS sector 42 (Wholesale Trade)<br>% jobs in NAICS sector 44-45 (Retail Trade)<br>% jobs in NAICS sector 48-49 (Transportation and Warehousing)<br>% jobs in NAICS sector 51 (Information)<br>% jobs in NAICS sector 52 (Finance and Insurance)<br>% jobs in NAICS sector 53 (Real Estate and Rental and Leasing)<br>% jobs in NAICS sector 54 (Professional, Scientific, and Technical Services)<br>% jobs in NAICS sector 55 (Management of Companies and Enterprises)<br>% jobs in NAICS sector 56 (Administrative and Support and Waste Management and Remediation Services)<br>% jobs in NAICS sector 61 (Educational Services)<br>% jobs in NAICS sector 62 (Health Care and Social Assistance)<br>% jobs in NAICS sector 71 (Arts, Entertainment, and Recreation)<br>% jobs in NAICS sector 72 (Accommodation and Food Services)<br>% jobs in NAICS sector 81 (Other Services [except Public Administration]) |
| <b>WAC</b> | % jobs for workers age 29 or younger<br>% jobs for workers age 30 to 54<br>% jobs with earnings \$1250/month or less<br>% jobs with earnings \$1251/month to \$3333/month<br>% jobs in NAICS sector 11 (Agriculture, Forestry, Fishing and Hunting)<br>% jobs in NAICS sector 21 (Mining, Quarrying, and Oil and Gas Extraction)<br>% jobs in NAICS sector 22 (Utilities)<br>% jobs in NAICS sector 23 (Construction)<br>% jobs in NAICS sector 31-33 (Manufacturing)<br>% jobs in NAICS sector 42 (Wholesale Trade)<br>% jobs in NAICS sector 44-45 (Retail Trade)<br>% jobs in NAICS sector 48-49 (Transportation and Warehousing)<br>% jobs in NAICS sector 51 (Information)<br>% jobs in NAICS sector 52 (Finance and Insurance)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |

Continued on next page

Table A9 – continued from previous page

| Source        | Variables                                                                                            |
|---------------|------------------------------------------------------------------------------------------------------|
|               | % jobs in NAICS sector 53 (Real Estate and Rental and Leasing)                                       |
|               | % jobs in NAICS sector 54 (Professional, Scientific, and Technical Services)                         |
|               | % jobs in NAICS sector 55 (Management of Companies and Enterprises)                                  |
|               | % jobs in NAICS sector 56 (Administrative and Support and Waste Management and Remediation Services) |
|               | % jobs in NAICS sector 61 (Educational Services)                                                     |
|               | % jobs in NAICS sector 62 (Health Care and Social Assistance)                                        |
|               | % jobs in NAICS sector 71 (Arts, Entertainment, and Recreation)                                      |
|               | % jobs in NAICS sector 72 (Accommodation and Food Services)                                          |
|               | % jobs in NAICS sector 81 (Other Services [except Public Administration])                            |
| <b>ZBP</b>    | Employment density in the workplace Tract                                                            |
|               | % high-tech companies with more than 100 employees                                                   |
|               | % high-tech companies with more than 250 employees                                                   |
|               | % high-tech companies with more than 500 employees                                                   |
|               | % high-tech companies with more than 1000 employees                                                  |
|               | % high-tech jobs = high-tech jobs / employment                                                       |
|               | % high-tech jobs in big companies = high-tech jobs in companies with 250+ employees / high-tech jobs |
| <b>Census</b> | Population density in the residential Tract                                                          |
|               | log of median rent adjusted to 2010                                                                  |
|               | log of median home value adjusted to 2010                                                            |
|               | log of median household income                                                                       |
|               | log of per capita income                                                                             |
|               | log of median earning adjusted to 2010 during past 12 months of low-skill group                      |
|               | log of median earning adjusted to 2010 during past 12 months of high-skill group                     |
|               | % population with at least a four-year college degree                                                |
|               | % population in poverty                                                                              |
|               | % unemployment                                                                                       |
|               | % owner-occupied housing units                                                                       |
|               | % vacant housing units                                                                               |
|               | % housing units in multi-unit structures                                                             |
|               | % household heads moved in less than 10 years                                                        |
|               | % population in age 25-34                                                                            |
|               | % population in age 35-44                                                                            |
|               | % population in age 45-54                                                                            |
|               | % population in age 55-64                                                                            |
|               | % population in age 65+                                                                              |
|               | % labor force in home tract                                                                          |
|               | log of weighted average building age of all the existent housing unit                                |
|               | % housing units built in last 11 to 20 years                                                         |
|               | % housing units built in last 21 to 30 years                                                         |
|               | % housing units built in last 31 to 40 years                                                         |
|               | % housing units built more than 40 years                                                             |
|               | log Distance                                                                                         |
|               | log Distance <sup>2</sup>                                                                            |

Notes: This tables shows all the candidate explanatory variables in the model of commute pattern by skill. They are from LODES OD, WAC, RAC, Census/ACS and ZBP. The construction of the model is presented in [C.5.1](#).



Table A10: Comparison of Predictive Models of Commute Pattern by Skill

|                | Full   |             |        |        |        | LODES  |             |        |        |        |
|----------------|--------|-------------|--------|--------|--------|--------|-------------|--------|--------|--------|
|                | Lasso  | Elastic Net | Ridge  | RF     | XGB    | Lasso  | Elastic Net | Ridge  | RF     | XGB    |
| No. Parameters | 23     | 23          | 87     | 87     | 87     | 28     | 27          | 55     | 55     | 55     |
| Data Source    |        |             |        |        |        |        |             |        |        |        |
| OD             | Y      | Y           | Y      | Y      | Y      | Y      | Y           | Y      | Y      | Y      |
| WAC            | Y      | Y           | Y      | Y      | Y      | Y      | Y           | Y      | Y      | Y      |
| RAC            | Y      | Y           | Y      | Y      | Y      | Y      | Y           | Y      | Y      | Y      |
| Census         | Y      | Y           | Y      | Y      | Y      |        |             |        |        |        |
| ZBP            | Y      | Y           | Y      | Y      | Y      |        |             |        |        |        |
| Observations   | 12,359 | 12,359      | 12,359 | 12,359 | 12,359 | 12,359 | 12,359      | 12,359 | 12,359 | 12,359 |
| Pseudo $R^2$   | 0.127  | 0.127       | 0.132  |        |        | 0.111  | 0.109       | 0.114  |        |        |
| Training RMSE  | 0.263  | 0.262       | 0.258  | 0.196  | 0.085  | 0.258  | 0.257       | 0.256  | 0.198  | 0.086  |
| Test RMSE      | 0.233  | 0.233       | 0.236  | 0.235  | 0.270  | 0.232  | 0.231       | 0.235  | 0.234  | 0.278  |

Notes: This table shows the performance of different models for predicting  $S_{i,j}^s$ , the share of workers who live in  $i$  and work in  $j$ , and belong to skill group  $s$ . Column 1 to column 5 are the models using the full set of predictors from LODES OD, WAC, RAC, Census/ACS and ZBP; Column 6 to 10 are the models using predictors only from LODES OD, WAC and RAC for robustness checks. Lasso, Elastic Net and Ridge are Logistic models with a penalized term. RF stands for Random Forest. XGB stands for XGBoost. For penalized regressions like Lasso, we use 10-fold cross-validation on the test sample to select the optimal penalty parameter. When we evaluate predictive accuracy using RMSE, we exclude the Tract pairs where all individuals belong to the same skill group.

Table A11: Estimated Parameters of Model for Commute Pattern by Skill

| Source | Variable                                                                                             | Coefficient |
|--------|------------------------------------------------------------------------------------------------------|-------------|
|        | Constant                                                                                             | -0.385      |
| OD     | % Jobs of workers age 29 or younger                                                                  | -1.089      |
|        | % Jobs with earnings \$1250/month or less                                                            | -0.521      |
|        | % Jobs with earnings \$1251/month to \$3333/month                                                    | -0.739      |
|        | % Jobs in Goods Producing industry sectors                                                           | -1.144      |
|        | % Jobs in Trade, Transportation, and Utilities industry sector                                       | -0.340      |
| RAC    | % Jobs with earnings \$1251/month to \$3333/month                                                    | -0.591      |
|        | % Jobs in NAICS sector 48-49 (Transportation and Warehousing)                                        | 2.079       |
|        | % Jobs in NAICS sector 53 (Real Estate and Rental and Leasing)                                       | 1.574       |
|        | % Jobs in NAICS sector 56 (Administrative and Support and Waste Management and Remediation Services) | -3.506      |
|        | % Jobs in NAICS sector 62 (Health Care and Social Assistance)                                        | -0.136      |
| WAC    | % Jobs in NAICS sector 23 (Construction)                                                             | -0.648      |
|        | % Jobs in NAICS sector 52 (Finance and Insurance)                                                    | -0.829      |
|        | % Jobs in NAICS sector 53 (Real Estate and Rental and Leasing)                                       | -1.102      |
|        | % Jobs in NAICS sector 55 (Management of Companies and Enterprises)                                  | 1.513       |
| ZBP    | Employment density in the workplace Tract                                                            | 0.865       |
|        | % High-tech jobs = high-tech jobs / employment                                                       | 0.236       |
|        | % High-tech jobs in big companies = high-tech jobs in companies with 250+ employees / high-tech jobs | 0.0302      |
| Census | % Population with at least a four-year college degree                                                | 3.119       |
|        | % Unemployment                                                                                       | 0.226       |
|        | % Population age 55-64                                                                               | 1.396       |
|        | % Housing units built in last 11 to 20 years                                                         | 0.647       |
|        | % Housing units built in last 21 to 30 years                                                         | 0.129       |

Notes: This tables shows the estimated parameters of the model for commute pattern by skill. The candidate explanatory variables are from the full set of predictors from LODES OD, WAC, RAC, Census/ACS and ZBP, as listed in Table A9.

Table A12: Commuting Elasticity

|                                                        | High-skilled             |                          | Low-skilled              |                          |
|--------------------------------------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
|                                                        | $t = 1$<br>(1)           | $t = 2$<br>(2)           | $t = 1$<br>(3)           | $t = 2$<br>(4)           |
| Commuting Elasticity ( $-\frac{\lambda^s}{\sigma^s}$ ) | -0.0233***<br>(0.000207) | -0.0217***<br>(0.000199) | -0.0232***<br>(0.000207) | -0.0212***<br>(0.000199) |
| Observations                                           | 1,049,113                | 1,074,080                | 1,049,113                | 1,074,080                |
| R-squared                                              | 0.716                    | 0.729                    | 0.710                    | 0.722                    |
| Home FE                                                | Yes                      | Yes                      | Yes                      | Yes                      |
| Work FE                                                | Yes                      | Yes                      | Yes                      | Yes                      |
| Entry Zone FE                                          | Yes                      | Yes                      | Yes                      | Yes                      |

Notes: This table lists the estimated commuting elasticity for period 1 and 2, and for high-skilled and low-skilled workers, respectively. All regressions includes fixed effects for home ZIP Code, Work ZIP Code and firm entry zone. \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ .

Table A13: Annual Welfare Change Due to Firm Entry by Initial Location in  $\tau = -5$ 

|                              | High-skilled    |                  | Low-skilled    |                  |
|------------------------------|-----------------|------------------|----------------|------------------|
|                              | Homeowner       | Renter           | Homeowner      | Renter           |
| <b>Total: All incumbents</b> | 110.5<br>0.21%  | -32<br>-0.06%    | 20.7<br>0.07%  | -55.2<br>-0.20%  |
| Within 10 mins               | 219.0<br>0.42%  | -149.4<br>-0.28% | 33.7<br>0.12%  | -180.4<br>-0.62% |
| Within 10–20 mins            | 175.6<br>0.34%  | -73.2<br>-0.14%  | 23.8<br>0.08%  | -123.3<br>-0.43% |
| Within 20–30 mins            | 122.9<br>0.24%  | -18.0<br>-0.03%  | 12.0<br>0.04%  | -70.8<br>-0.25%  |
| Within 30 mins–50 miles      | -0.6<br>-0.001% | -5.8<br>-0.01%   | -7.4<br>-0.03% | -10.2<br>-0.04%  |

Notes: For each group of incumbents by their initial locations, the top row reports the annual changes in expected utility in real 2010 dollars for a representative high-skill firm entry, 5 years after entry. The bottom row reports the annual changes in expected utility as percent changes in incumbents' average wages in period 1, which is one year prior to the firm entry.

Table A14: Welfare Change: Movers vs. Stayers

|                  | High-skilled |         | Low-skilled |         | Total  |
|------------------|--------------|---------|-------------|---------|--------|
|                  | Homeowner    | Renter  | Homeowner   | Renter  |        |
| Move to Diff Zip | 107.7        | -36.6   | 6.5         | -71.2   | 12.7   |
| Share            | [27.3%]      | [8.66%] | [41.1%]     | [22.9%] | [100%] |
| Stay             | 72.1         | -31.5   | 7.0         | -49.5   | 4.4    |
| Share            | [25.1%]      | [9.26%] | [40.8%]     | [24.8%] | [100%] |

Notes: The first row reports the annual average change in a mover's expected utility in real 2010 dollars across 129 firm entries, five years after the firm entry announcement. The third row reports that for the average stayer. These welfare numbers can be interpreted as the welfare change for a representative high-skilled firm entry in our sample. In the second and fourth rows, we report in brackets the share of incumbent workers who belong to each skill and homeownership group, for movers and stayers, respectively.

Table A15: Neighborhood Amenity Characteristics

| Variable Name        | Definition of Variable                                                                                                                                                                        | Data Source                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | NAICS Codes                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|----------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Log Retail Amenities | Log total number of retail amenities per 1000 people, including supermarkets, grocery stores, convenience stores, apparel stores, etc.                                                        | The source of establishments is ZIP Codes Business Pattern data. We select the related sub-industries based on NAICS and sum them to get the total number of establishments in the industry. Because there are gaps of these measurements around 1997, we use 1998-2002 NAICS data to linearly extrapolate values from 1994-1997 if the data is missing. The source of total population at ZIP Codes level is ACS 5-year Census. In the end, we divide the total number of establishments to the total population at ZIP Codes level and take the logarithm. | "445110", Supermarkets and Other Grocery (except Convenience) Stores;<br>"445120", Convenience Stores;<br>"445310", Beer, Wine, and Liquor Stores;<br>"446110", Pharmacies and Drug Stores;<br>"446120", Cosmetics, Beauty Supplies, and Perfume Stores;<br>"446130", Optical Goods Stores;<br>"446191", Food (Health) Supplement Stores;<br>"446199", All Other Health and Personal Care Stores;<br>"448110", Men's Clothing Stores;<br>"448120", Women's Clothing Stores;<br>"448130", Children's and Infants' Clothing Stores;<br>"448140", Family Clothing Stores;<br>"448150", Clothing Accessories Stores;<br>"448190", Other Clothing Stores;<br>"448310", Jewelry Stores;<br>"451110", Sporting Goods Stores;<br>"451120", Hobby, Toy, and Game Stores;<br>"451140", Musical Instrument and Supplies Stores.<br>"451211", Book Stores;<br>"451212", News Dealers and Newsstands;<br>"452110", Department Stores (except Discount Department Stores);<br>"452910", Warehouse Clubs and Supercenters |
| Log Private Services | Log total number of personal care services places per 1000 people, including barber shop, salon, laundry services and other personal care services                                            | Same source as above                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | "812111", Barber Shops;<br>"812112", Beauty Salons;<br>"812113", Nail Salons;<br>"812191", Diet and Weight Reducing Centers;<br>"812199", Other Personal Care Services;<br>"812320", Drycleaning and Laundry Services (except Coin-Operated).                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| Log Restaurant       | Log total number of restaurants per 1000 people, including full-service restaurants, limited-service restaurants cafeterias, grill buffets, and buffets snack and nonalcoholic beverage bars. | Same source as above                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | "722110", Full-service restaurants<br>"722211", Limited-service restaurants<br>"722212", Cafeterias, grill buffets, and buffets<br>"722213", Snack and nonalcoholic beverage bars                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| Log Nightlife        | Log total number of drinking places per 1000 people, which mainly offers alcoholic beverages.                                                                                                 | Same source as above                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | "722410", Drinking Places (Alcoholic Beverages)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| Log Recreation       | Log total number of recreation places per 1000 people, including theaters, sports centers, museums and amusement parks.                                                                       | Same source as above                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | "512131", Motion Picture Theater;<br>"711110", Theater Companies and Dinner Theaters;<br>"711211", Sports Teams and Clubs;<br>"711212", Racetracks;<br>"711219", Other Spectator Sports;<br>"713950", Bowling Centers;<br>"713910", Golf Courses and Country Clubs;<br>"713940", Fitness and Recreational Sports Centers;<br>"712110", Museums;<br>"713110", Amusement and Theme Parks;<br>"713120", Amusement Arcades;<br>"713990", All Other Amusement and Recreation Industries.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |

Notes: The Census population data is retrieved from NHGIS. The ZIP Code Business Pattern data are available from 1994 onwards for 2- through 6-digit NAICS code level. There are 3 versions for the NAICS codes released in 1997, 2002 and 2012 respectively. Though the 6-digit code of each sub-industry may vary in different versions, the name of sub-industries will stay the same. We choose to present the latest version of 6-digit code in the NAICS Codes column.

Table A16: Neighborhood Demographic Characteristics for Constructing the Propensity Score Model

| Variable Name                                         | Definition of Variable                                                                                                                     | Data Source                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|-------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Log per capita income                                 | log per capita annual income of residents in successive rings in 1990                                                                      | We derive Tract-level per capita income from 1990 Census downloaded via NHGIS. Tracts are in 2010 boundaries and per capital income is adjusted to 2010 real dollars. Using the 2010 ZCTA to Census Tract Relationship File from <a href="#">Census</a> , we crosswalk them to the 2010 ZCTA level. We use the percentage of ZCTA's population in each Tract (ZPOPPCT) in 2010 as weights. Then we calculate the weighted average measures in successive rings using population as weights.                |
| Log median household income                           | log median annual income of households in successive rings in 1990                                                                         | Same source as above                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| % Population age under 19                             | share of population aged 19 or below in successive rings in 1990                                                                           | Same source as above                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| % Population age 20-44                                | share of population aged from 20 to 44 in successive rings in 1990                                                                         | Same source as above                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| % Population age 45-64                                | share of population aged from 45 to 64 in successive rings in 1990                                                                         | Same source as above                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| % Population age 65+                                  | share of population aged 65 or above in successive rings in 1990                                                                           | Same source as above                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| % Non-hispanic white                                  | share of population of white race (not hispanic orgin) in successive rings in 1990                                                         | Same source as above                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| % Population with at least a four-year college degree | number of population with at least a four-year college degree divided by number of population aged 25 or above in successive rings in 1990 | Same source as above                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| % Poor                                                | share of population below the property line in successive rings in 1990                                                                    | Same source as above                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| % Unemployed                                          | share of unemployed population in successive rings in 1990                                                                                 | Same source as above                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| Population density                                    | population per unit area                                                                                                                   | Population is from the same source as above and total area is from 2010 U.S. Census Tract shape files downloaded from NHGIS. The unit is population per square meters.                                                                                                                                                                                                                                                                                                                                     |
| % Housing units which are renter-occupied             | share of renter-occupied housing units in successive rings in 1990                                                                         | We derive Tract-level per capita income from 1990 Census downloaded via NHGIS. Tracts are in 2010 boundaries and per capital income is adjusted to 2010 real dollars.Using the 2010 ZCTA to Census Tract Relationship File from <a href="#">Census</a> , we crosswalk them to the ZCTA level. We use the percentage of ZCTA's housing units in each Tract (ZHUPCT) in 2010 as weights. Then we calculate the weighted average measures in successive rings using number of housing units (ZHU) as weights. |
| % Housing units in multi-unit structures              | share of housing units in multi-unit structures in successive rings in 1990                                                                | Same source as above                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| % Housing units which are vacant                      | share of vacant housing units in successive rings in 1990                                                                                  | Same source as above                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| % Household heads moved into unit less than 10 year   | share of household heads who moved into unit less than 10 years in successive rings in 1990                                                | Same source as above                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| Log median rent                                       | log median rent of renter-occupied housing units in successive rings in 1990                                                               | Same source as above                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| Log median home value                                 | log median home value of housing units in successive rings in 1990                                                                         | Same source as above                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |

Notes: ZCTAs are defined by 2010 ZCTA Tabulation Areas (ZCTAs), and are time-invariant in their boundaries. For all the predictors in the propensity score model except “distance to the closest CBD”, we do not directly use the a ZCTA’s own characteristics. Instead, we use the average features of neighborhoods in successive rings around each ZCTA, defined as ZCTAs within 0–5, 5–10, 10–20, 20–30 miles away from each ZCTA. For a given ZCTA, we calculate the weighted average features of all the ZCTAs within each ring around it.

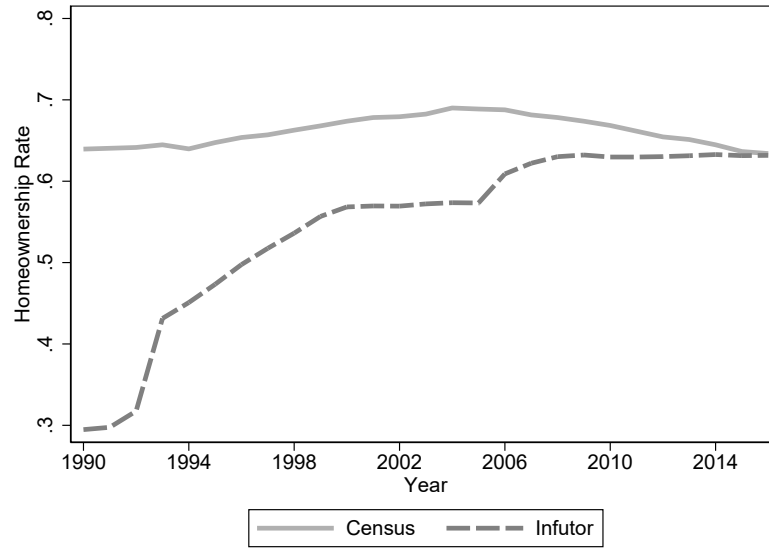
Table A17: Neighborhood Workplace Characteristics for Constructing the Propensity Score Model

| Variable Name                                 | Definition of Variable                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | Data Source                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|-----------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Log mean workplace wage                       | log mean annual wage of jobs in successive rings in 1990                                                                                                                                                                                                                                                                                                                                                                                                                                    | First, we derive the 2010 ZCTA-level workplace wage and number of jobs from ZIP Codes Business Patterns. The wage data are not disclosed in some ZCTAs, so we do extrapolation to fill in missing values using data from 1995 to 1999. Sometimes symbols indicating a particular employment size class (e.g.100-249 employees) are disclosed instead of the exact number of jobs. See the details in the <a href="#">documentation</a> . In these cases, we apply the middle of the class as a rough estimate. Then we calculate the weighted average measures in successive rings using ZCTA-level number of jobs as weights. |
| Job density in 12 Fama-French (FF) industries | number of jobs in each FF industry divided by total number of jobs in successive rings in 1994. 12 Fama-French industries, defined by 4-digit SIC codes, include the following: Consumer Nondurables, Consumer Durables, Manufacturing, Oil, Gas, and Coal Extraction and Products, Chemicals and Allied Products, Business Equipment, Telephone and Television Transmission, Utilities, Wholesale, Retail, and Some Services, Healthcare, Medical Equipment, and Drugs, Finance and Other. | We derive number of jobs by industry in each zipcode from ZIP Codes Business Patterns. 1994 is the earliest available year. Definitions of the FF industries can be downloaded from <a href="#">their website</a> .                                                                                                                                                                                                                                                                                                                                                                                                            |
| Hightech job density                          | number of jobs in hightech industries in successive rings divided by total number of jobs in hightech industries within 30 miles of firm entry in 1994.                                                                                                                                                                                                                                                                                                                                     | We use hightech industry definitions from <a href="#">National Science Foundation (2020)</a> and job numbers from ZCTA Business Pattern. Since NSF definitions are given by 4-digit 2002/2007 NAICS codes, we use concordances files from <a href="#">Census</a> to crosswalk them to SIC codes used in 1994.                                                                                                                                                                                                                                                                                                                  |
| Employment/population ratio                   | number of jobs in 1994 divided by population(number of residents) in 1990 in successive rings                                                                                                                                                                                                                                                                                                                                                                                               | We derive ZCTA-level number of jobs from ZIP Codes Business Patterns and population from 1990 Census. By aggregating the numbers to each bin, we get both the numerator and the denominator.                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| Distance to the closest CBD                   | distance from the centroid of each ZCTA to the closest CBD in miles                                                                                                                                                                                                                                                                                                                                                                                                                         | Definitions of CBDs are from <a href="#">Holian and Kahn (2015)</a> who used Google Earth to geocode the principal city of each of 366 MSAs by recording the latitude and longitude returned—the location where the map view centered—from a city name search.                                                                                                                                                                                                                                                                                                                                                                 |
| Distance to the closest CBD squared           | distance squared from each ZCTA to the closest CBD in miles                                                                                                                                                                                                                                                                                                                                                                                                                                 | Same source as before                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |

Notes: ZCTAs are defined by 2010 ZCTA Tabulation Areas (ZCTAs), and are time-invariant in their boundaries. For all the predictors in the propensity score model except “distance to the closest CBD”, we do not directly use the a ZCTA’s own characteristics. Instead, we use the average features of neighborhoods in successive rings around each ZCTA, defined as ZCTAs within 0–5, 5–10, 10–20, 20–30 miles away from each ZCTA. For a given ZCTA, we calculate the weighted average features of all the ZCTAs within each ring around it.

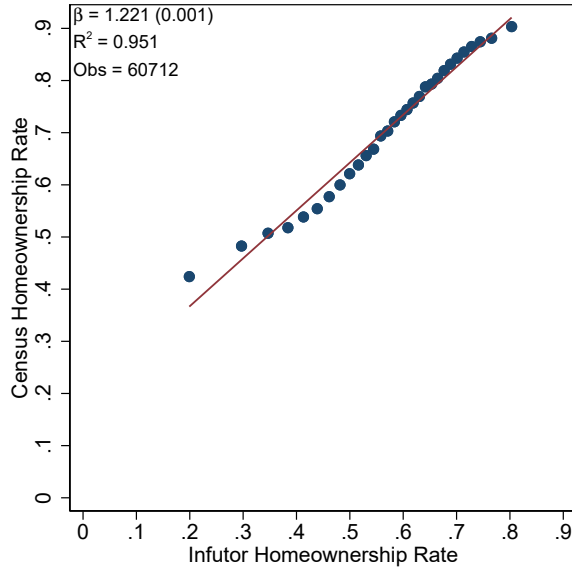
## B Additional Figures

Figure B1: Homeownership Rate Time Series

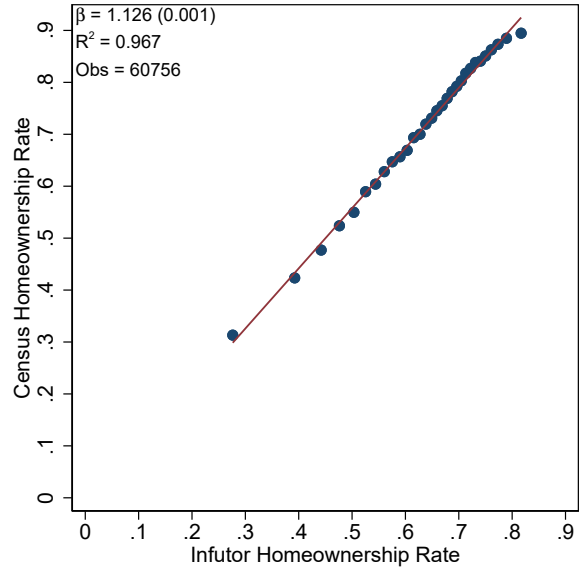


Notes: This figure compares the time series of homeownership rate at the national level from the U.S. Census Bureau and imputed from our Infutor-CoreLogic linked dataset. The Census homeownership rate is the annual average at the national level without seasonal adjustment. The Infutor homeownership rate is calculated using the full Infutor-CoreLogic linked dataset.

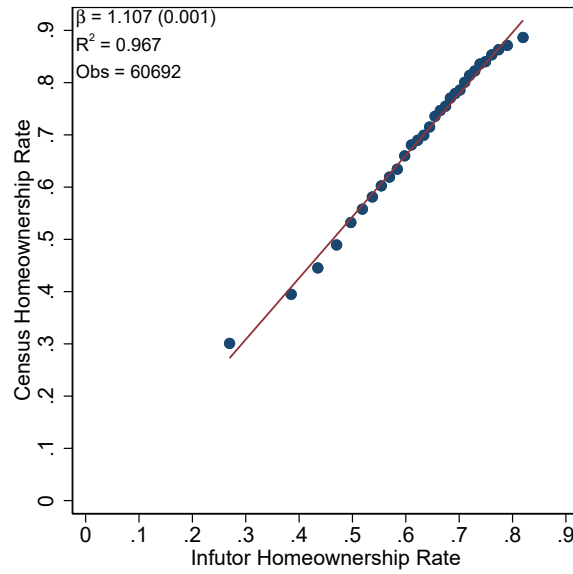
Figure B2: Validation for Homeownership Imputation



(a) 2000



(b) 2010

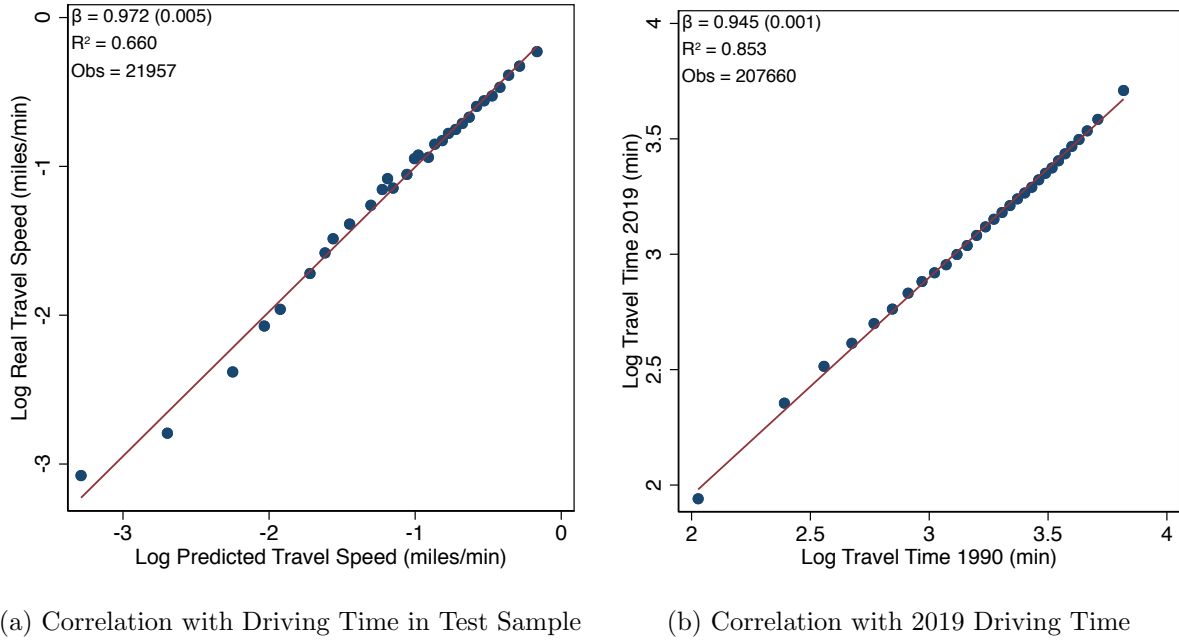


(c) 2015

Notes: This figure presents binscatter plots that validate the homeownership imputation at tract level. For three figures, we keep the tracts more than 10 percentile observations who can be imputed for a homeownership status. We divide the sample into 30 bins and plot the average value for each bin. The x-axis is the Infutor homeownership rate. The y-axis is the census homeownership rate. The regression is weighted by observations in each tract who can be imputed for a homeownership status and adjusted by the robust standard error. In Panel (a), we compare with the homeownership rate from the 2000 Decennial Census cross-walked to 2010 Census Tract boundaries. In Panel (b), we compare with the homeownership rate from the 2008–2012 5-year ACS. In Panel (c), we compare with the homeownership rate from the 2013–2017 5-year ACS.

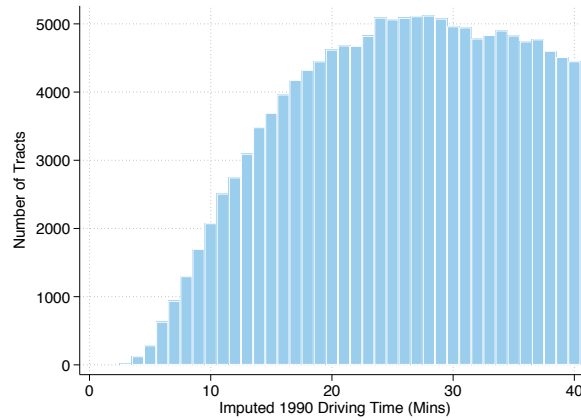


Figure B3: Travel Speed Model - Model Fit



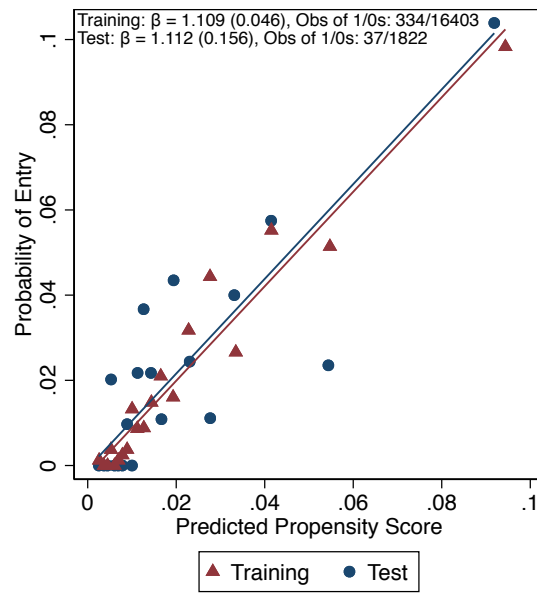
Notes: This figure present binscatter plots that validate the speed model. For both figures, we divide the sample into 30 bins and plot the average value for each bin. In Panel (a), we split the NHTS sample of trips into 80% training sample and 20% test sample. The x-axis is the log predicted travel speed from our speed model in the test sample. The y-axis is the log actual travel speed calculated from travel distance divided by recorded travel time in the survey. In Panel (b), the sample include all Census Tracts within 25 miles of each firm entry location. The x-axis is the log predicted travel speed from our speed model. The y-axis is the log of 2019 travel speed calculated using Google Maps Distance Matrix API.

Figure B4: 1990 Driving Time from Census tracts to Firm Site



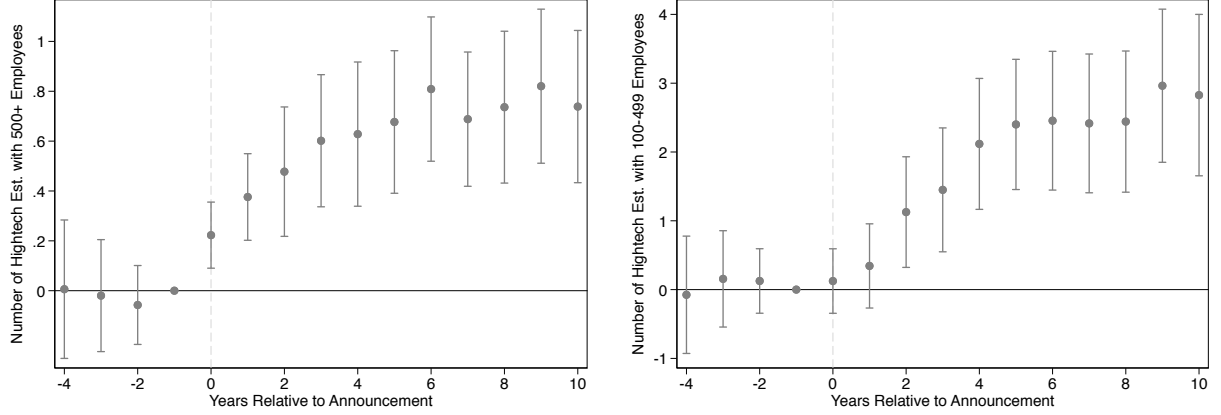
Notes: This figure shows a histogram of the number of Census tracts at various driving time bins from the entry sites in our sample. The x-axis is imputed driving time in 1990 during rush hour with traffic.

Figure B5: Propensity Score of Firm Entry – Model Fit

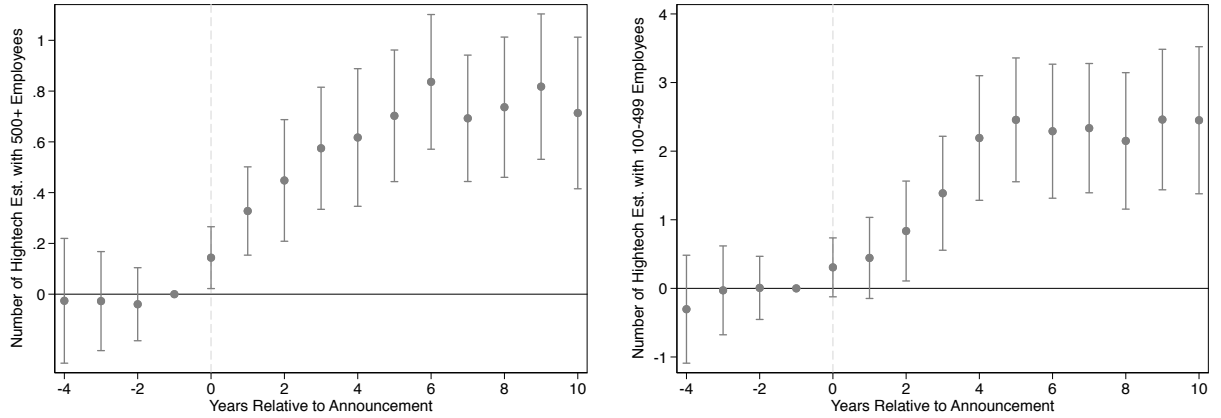


Notes: The figure shows the model fit for the propensity score of high-skilled firm entry. The sample of ZIP codes consists of all ZIP codes within 50 miles of each firm's entry location, and a ZIP code is in the treatment group if it was ever home to an entry in our sample. 90% of the 410 firms comprise the training data, and 10% of the 410 firms are held out as the testing data. We divide ZIP codes of the training and test data into 20 quantiles. The x-axis plots the average predicted probability of being home to an entry from the model in each quantile of ZIP codes, the y-axis plots the average probability of having an actual firm entry in each quantile.

Figure B6: Robustness Check: Effect on High-skilled Job Growth



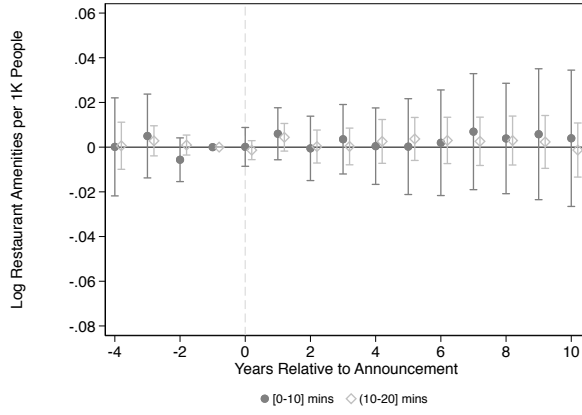
(a) High-tech Establishments with 500+ Employees, Within County Comparison (b) High-tech Establishments with 100-499 Employees, Within County Comparison



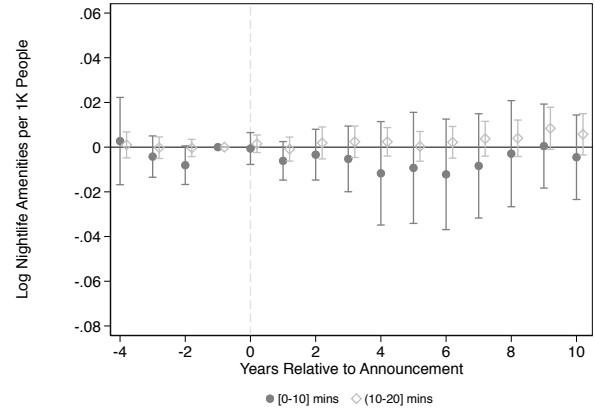
(c) High-tech Establishments with 500+ Employees, P-score Model with Decadal Changes (d) High-tech Establishments with 100-499 Employees, P-score Model with Decadal Changes

Notes: This figure uses event study to validate our propensity score model using equation (30). It presents the robustness checks for our baseline results in Figure 3. Table A6 also summarizes the event study coefficients in the post-entry periods. In Panels (a) and (b), we compare actual sites that received a firm entry in our sample with top potential sites that within the same county of the entry location. This further controls for potential unobserved differences between treated and control sites that could differ across counties within the same metropolitan area, such as differences in tax incentives. In Panels (c) and (d), we add decadal changes of neighborhood demographic and workplace characteristics to our p-score model for predicting potential sites of high-skilled firm entry. This further controls for dynamics in neighborhood characteristics that could affect firms' location decisions. The outcome variable is the number of high-tech establishments with 500+ employees at ZIP codes level in Panels (a) and (c), and the number of high-tech establishments with 100-499 employees at ZIP codes level in Panels (b) and (d). The data source is ZIP Code Business Patterns. The treatment group are ZIP Codes that actually received a firm entry from our sample of 391 firm entries. The control group are ZIP Codes that did not receive an entry in our sample, but are among the top quintile of ZIP Codes within 50 miles of the actual entry location that have the highest predicted probability of receiving a high-skilled firm entry. 95% confidence intervals are provided, where standard errors are clustered at the metropolitan area (50-mile ring around the actual entry location) level.

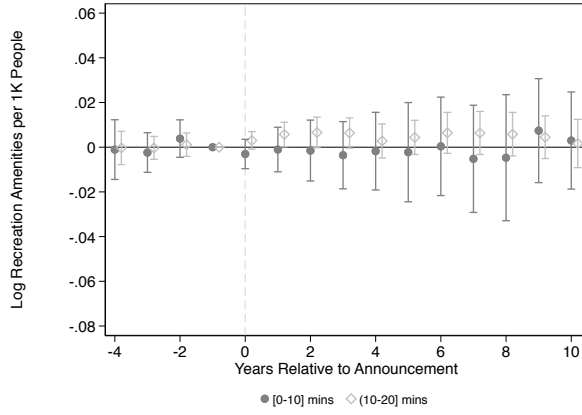
Figure B7: Dynamic Difference In Difference of Neighborhood Amenity Outcomes



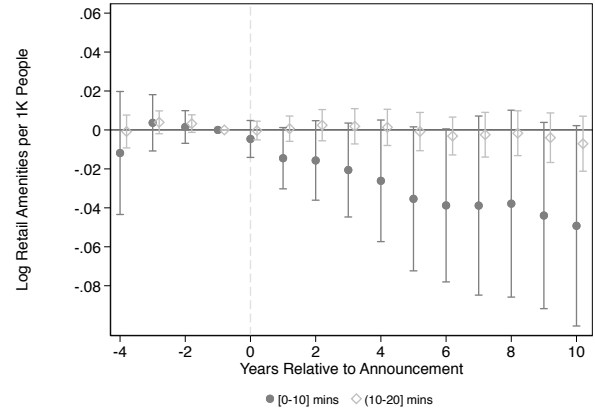
(a) Restaurant



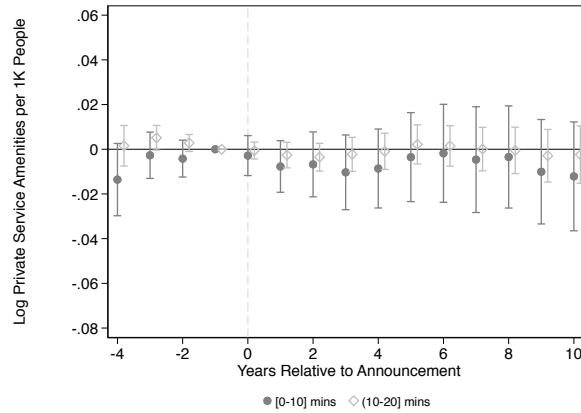
(b) Nightlife



(c) Recreation



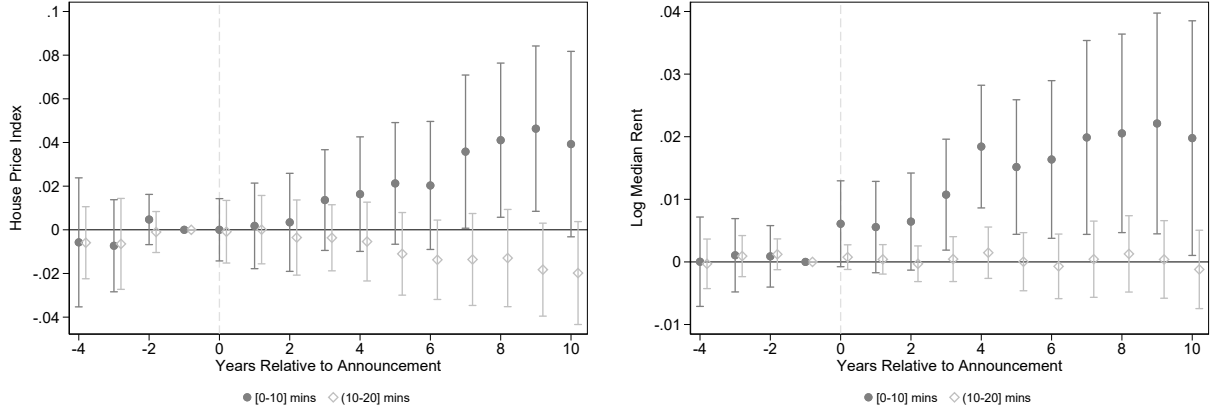
(d) Retail



(e) Private Service

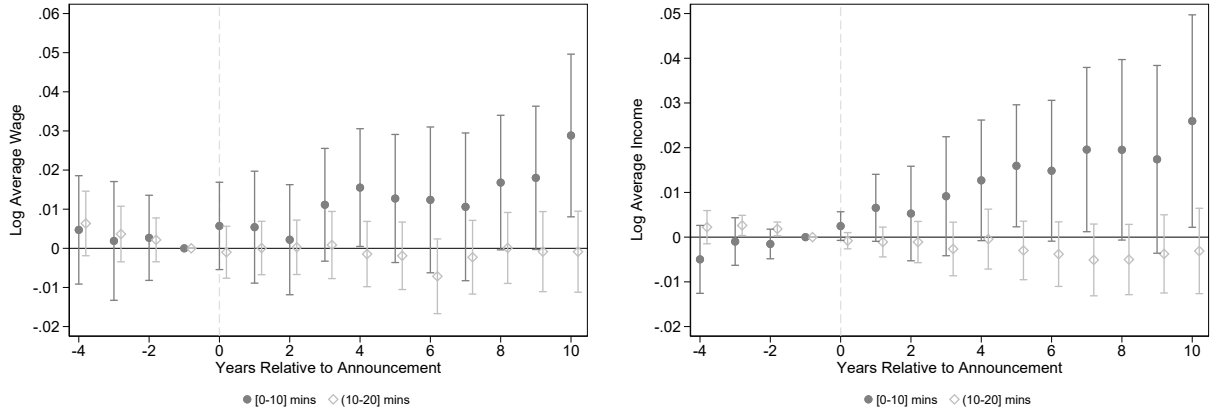
Notes: This figure presents the coefficients of treatment effect for the event study for local amenity outcomes at ZIP Code level. 95% confidence intervals are provided, where standard errors are clustered at the entry zone level. The panel (a) presents the treatment effect of log number of restaurants per 1000 people. The panel (b) presents the treatment effect of log number of nightlife amenities per 1000 people. The panel (c) presents the treatment effect of log number of recreation amenities per 1000 people. The panel (d) presents the treatment effect of log number of retail amenities per 1000 people. The panel (e) presents the treatment effect of log number of private service amenities per 1000 people. Table A15 documents how we construct measures of these amenities.

Figure B8: Robustness Check: Event Study for the Effects of High-skilled Firm Entries on Neighborhood Outcomes



(a) House Price Index in Residential Census Tract

(b) Median Rent in Residential Census Tract



(c) Workers' Avg. Income in Workplace ZIP Code

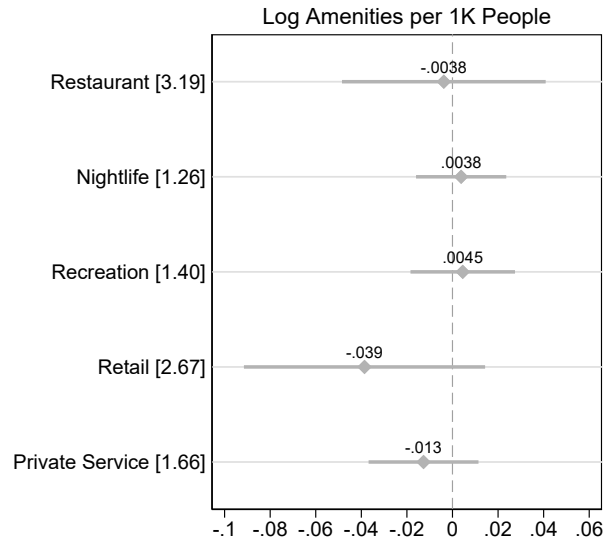
(d) Residents' Avg. Income in Residential ZIP Code

Notes: This figure serves as the robustness check for the results in Figure 4. We additionally control for a neighborhood's proximity to ZIP Codes with high prior density of high-tech employment. The figure plots the event study coefficients of treatment effects on several neighborhood outcomes for neighborhoods in the inner ring within 10 mins from the entry and in the middle ring within 10–20 mins from the entry, relative to those in the outer rings 20–30 mins away from the entry site, before/after the entry announcement. The house price index at the Census Tract level is constructed using CoreLogic transactions records. Appendix C.3 documents how we construct this index. Log median rent at Census Tract level is from Censuses/ACS. Log average income in a workplace ZIP Code is computed using the average ZIP Code level Payroll data from ZIP Code Business Patterns. Log average income in a residential ZIP Code is computed using the average ZIP Code level adjusted gross income (AGI) from the IRS Statistics of Income Tax Stats. 95% confidence intervals are provided, where standard errors are clustered at the entry zone level.

Figure B9: Robustness Check: The Effects of High-skilled Firm Entries on Neighborhood Outcomes 6–10 Years Later



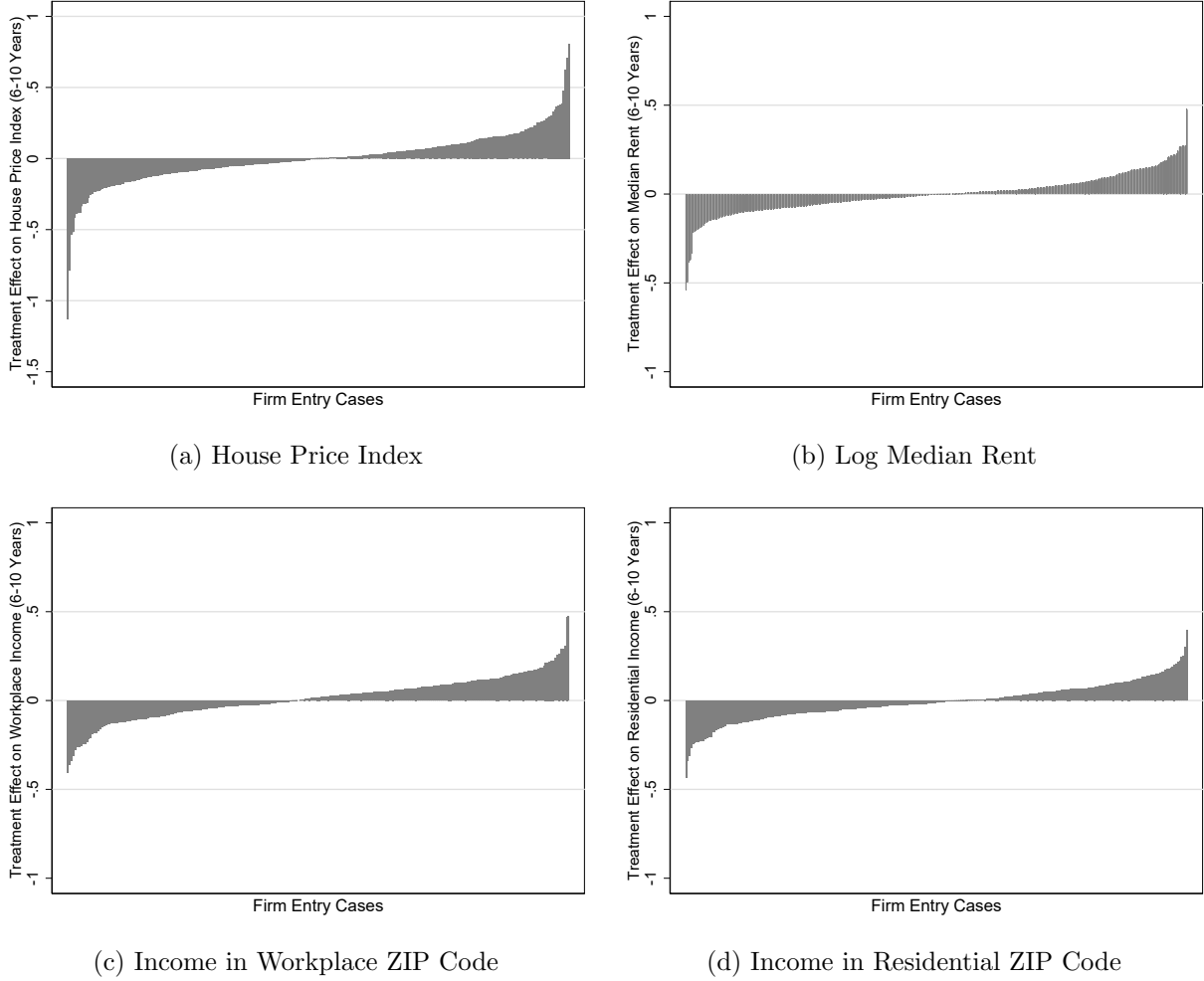
(a) Workplace and Residential Neighborhood Characteristics



(b) Neighborhood Amenities

Notes: This figure serves as the robustness check for the results in Figure 5. We additionally control for a neighborhood's proximity to ZIP Codes with high prior density of high-tech employment. The figure plots the average treatment effects on several neighborhood outcomes for neighborhoods in the inner ring within 10 mins from the firm entry, relative to those in the outer rings 20–30 mins away from the entry site, 6–10 years after the entry announcement. In Panel (a), the house price index at Census Tract level is constructed using CoreLogic transactions data. Appendix C.3 documents how we construct this index. Log median rent at Census Tract level is from Censuses/ACS. Log average income in a workplace ZIP Code is computed using the average ZIP Code level Payroll data from ZIP Code Business Patterns. Log average income in a residential ZIP Code is computed using the average ZIP Code level adjusted gross income (AGI) from the IRS Statistics of Income Tax Stats. Panel (b) plots the average treatment effects on ZIP Code level amenities 6–10 years later. Table A15 documents how we construct measures of these amenities. We present the mean of the dependent variables in the control group 6–10 years after the announcement in brackets. 95% confidence intervals are provided, where standard errors are clustered at the entry zone level.

Figure B10: Case-by-Case Treatment Effects on Neighborhood Outcomes



Notes: This figure shows the treatment effect on neighborhood outcomes for each individual firm entry. The treatment effect is estimated case by case for neighborhoods in the inner ring within 10 mins of driving time from the firm entry, relative to those in the outer rings 20–30 mins away from the entry site, 6–10 years after the announcement. House price index is computed using the Census Tract level hedonic house price indices constructed using CoreLogic transaction data. Appendix Section C.3 documents how we construct this index. Log median rent is computed from the decennial Census and ACS at the Census Tract level. Income in Workplace ZIP Code is computed using the average ZIP Code Payroll data from ZIP Code Business Patterns. Income in Residential ZIP Code is computed using the average residential ZIP Code income from the IRS Statistics of Income Tax Stats.

Figure B11: Dynamic Difference In Difference of Financial Outcomes

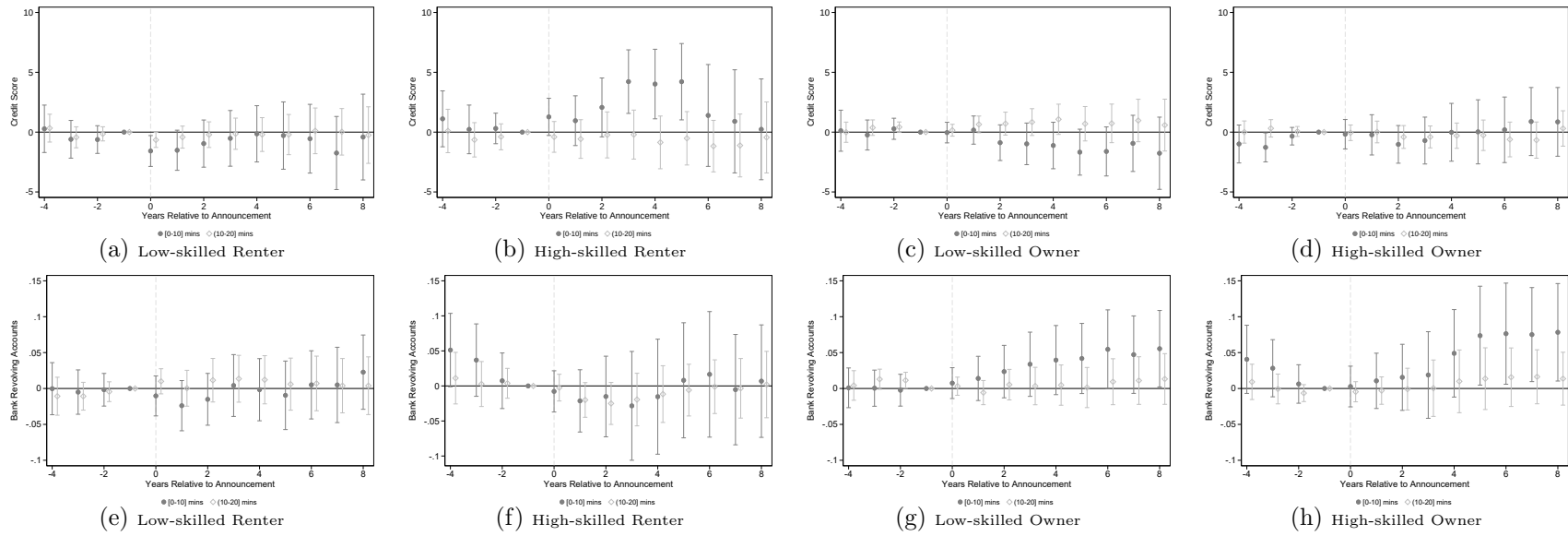




Figure B12: Dynamic Difference In Difference of Financial and Share Moved Outcomes

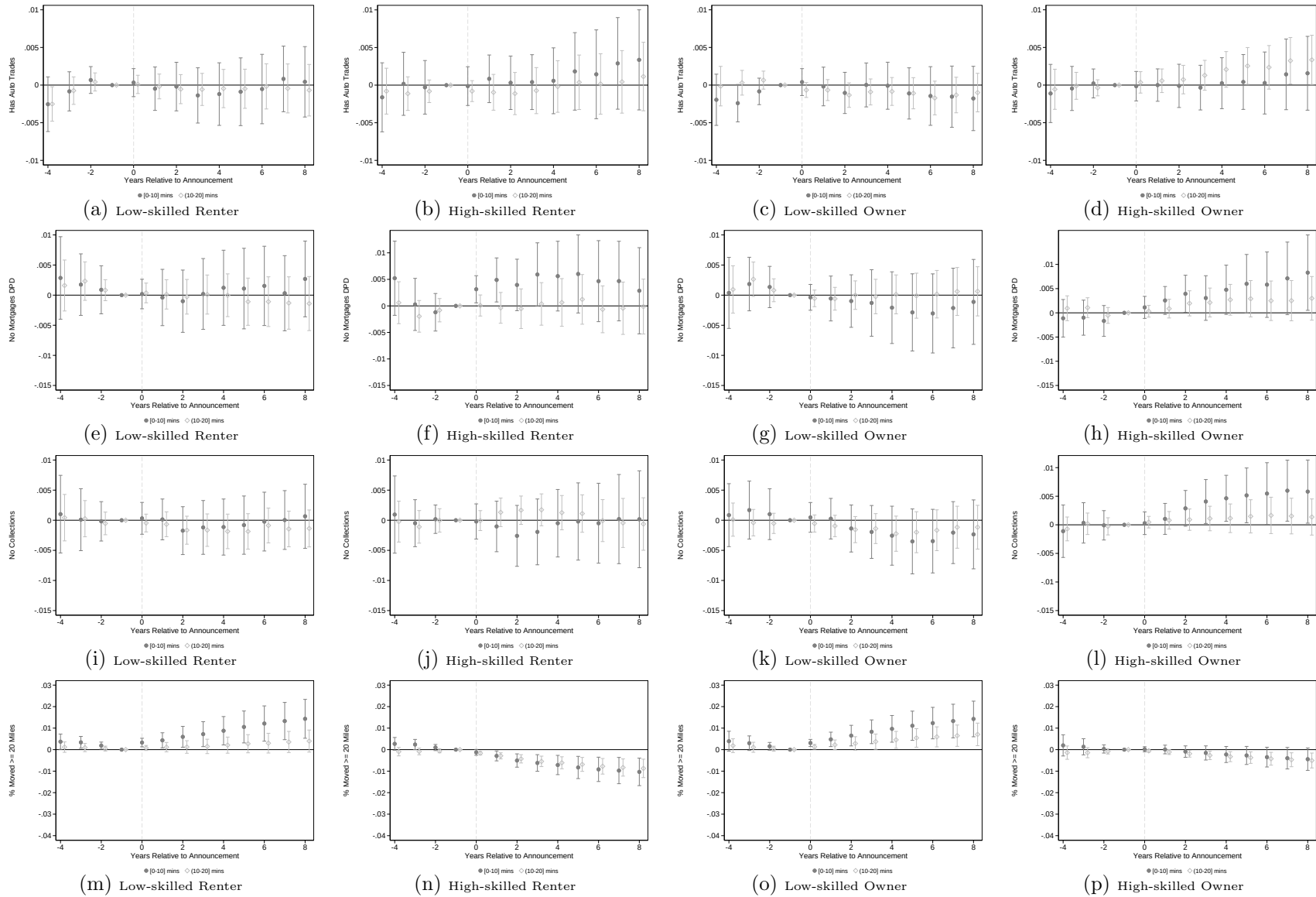


Figure B13: Dynamic Difference In Difference of Neighborhood Quality Outcomes

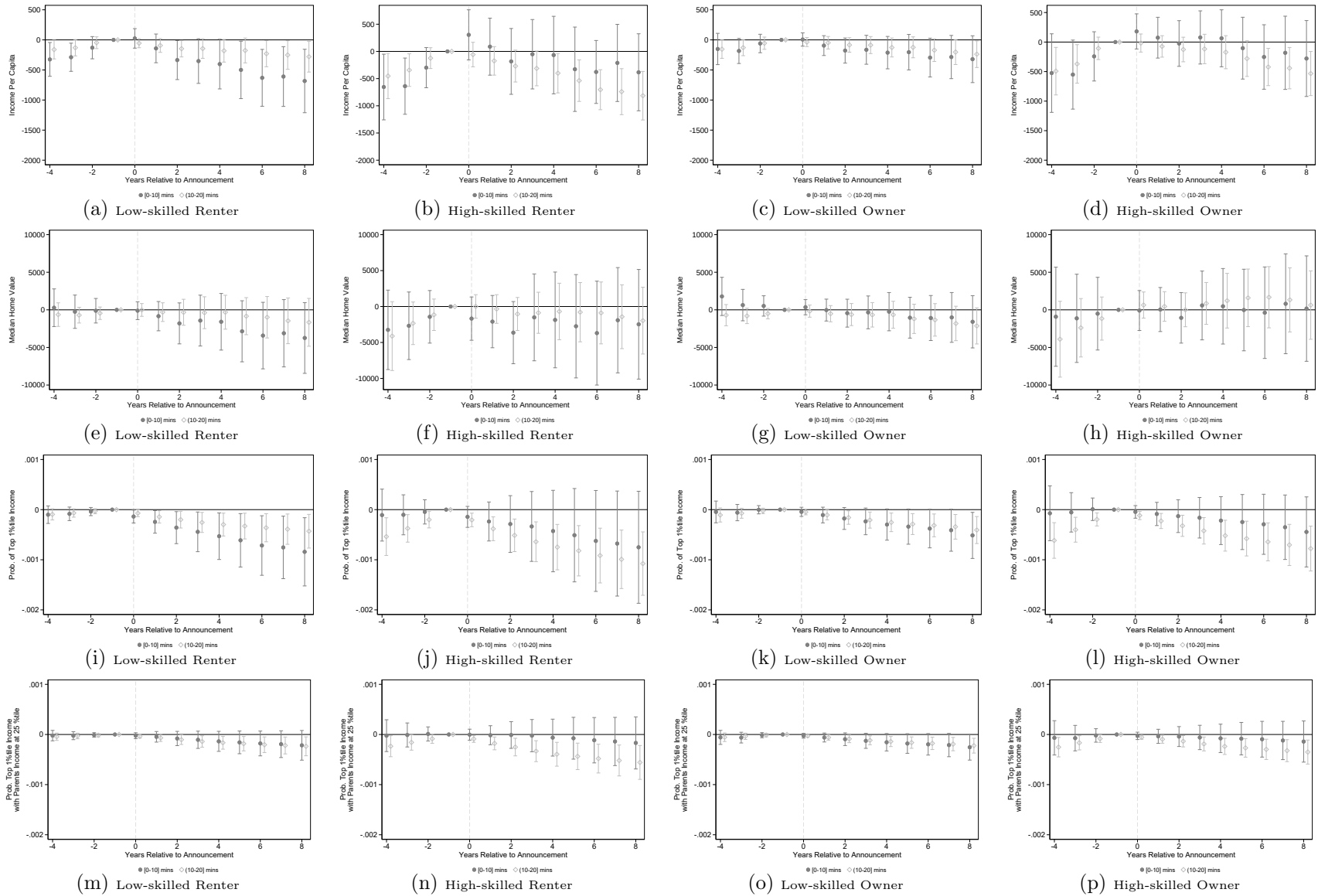
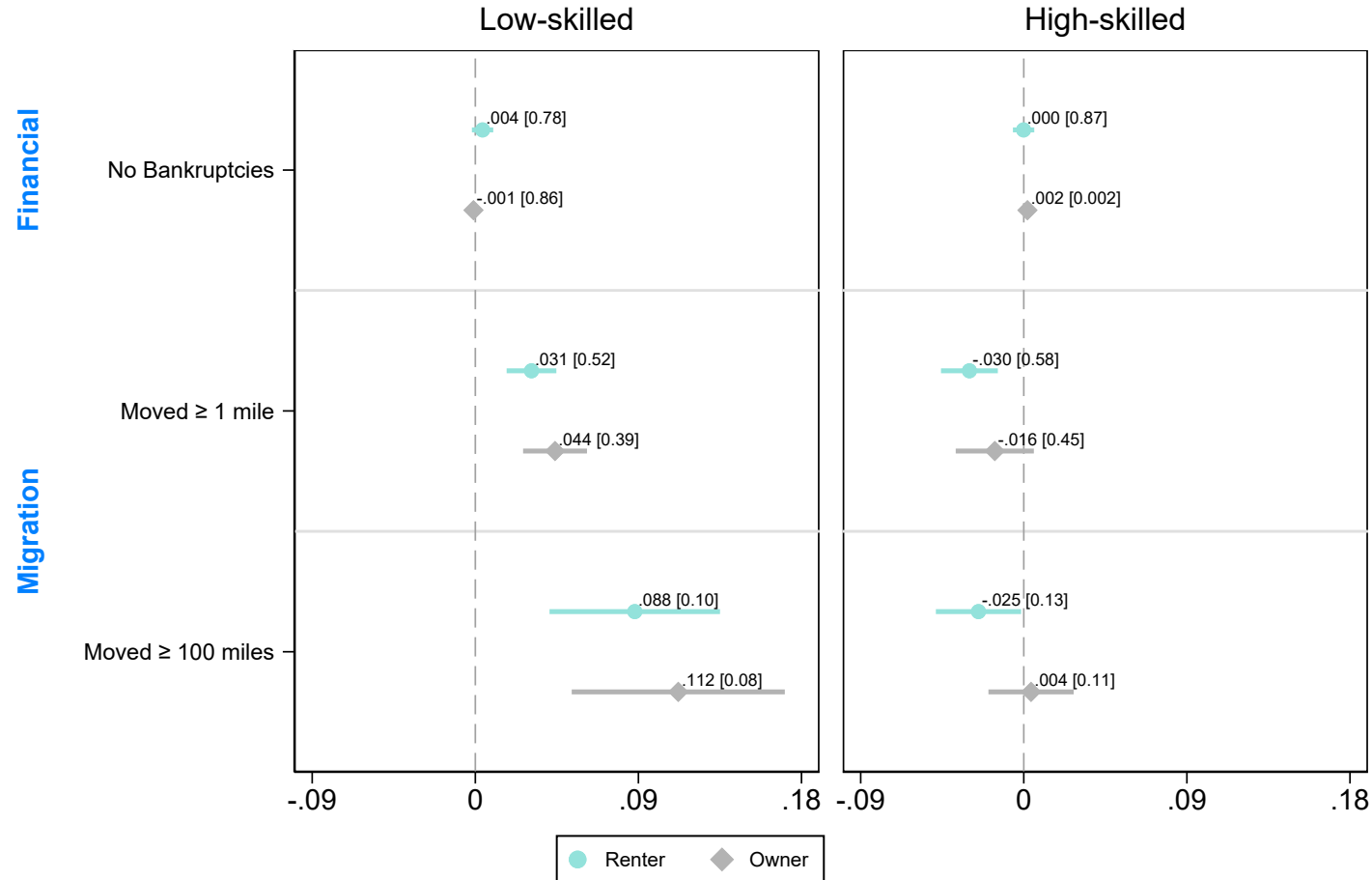
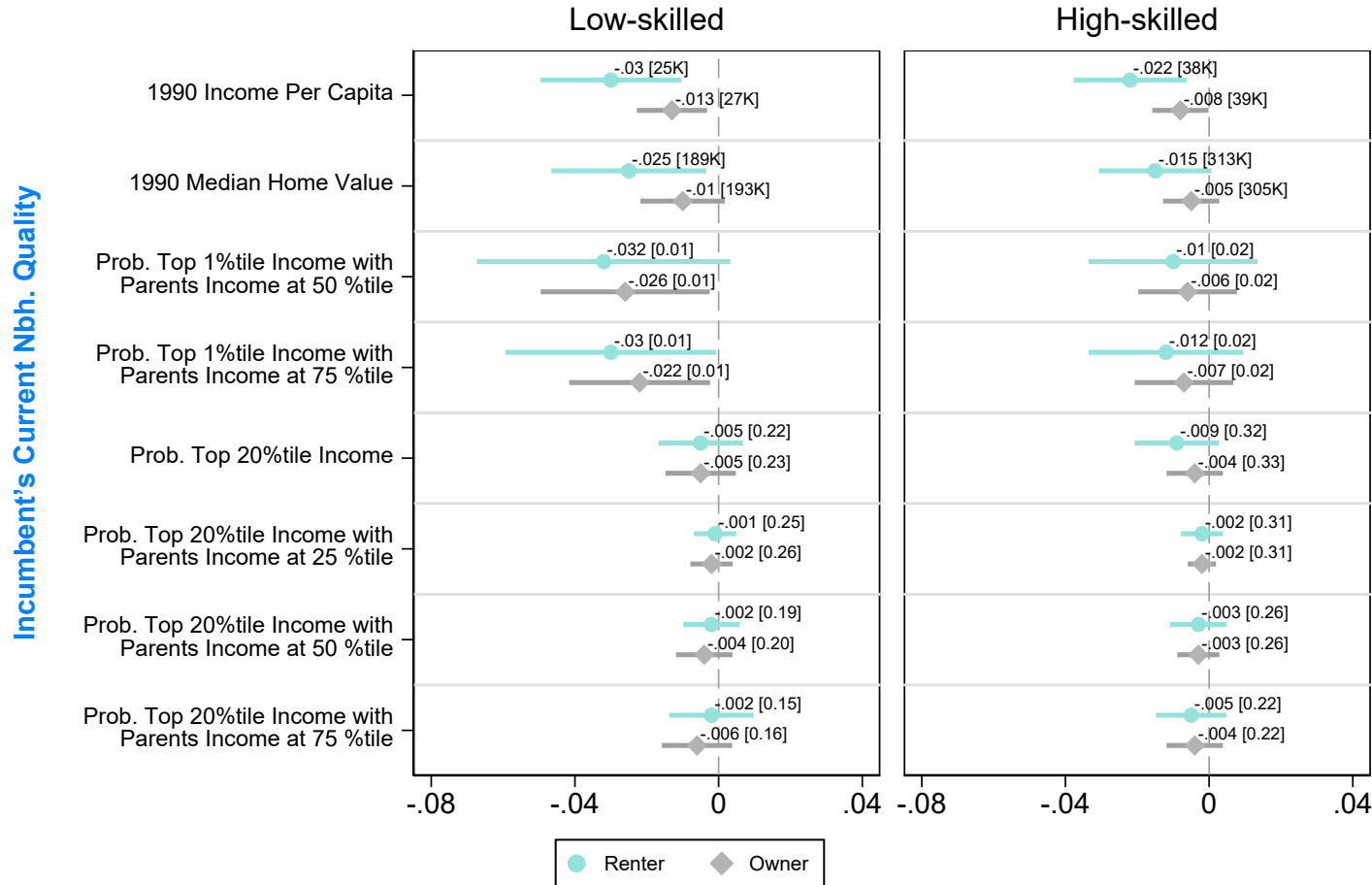


Figure B14: Financial and Share Moved Outcomes



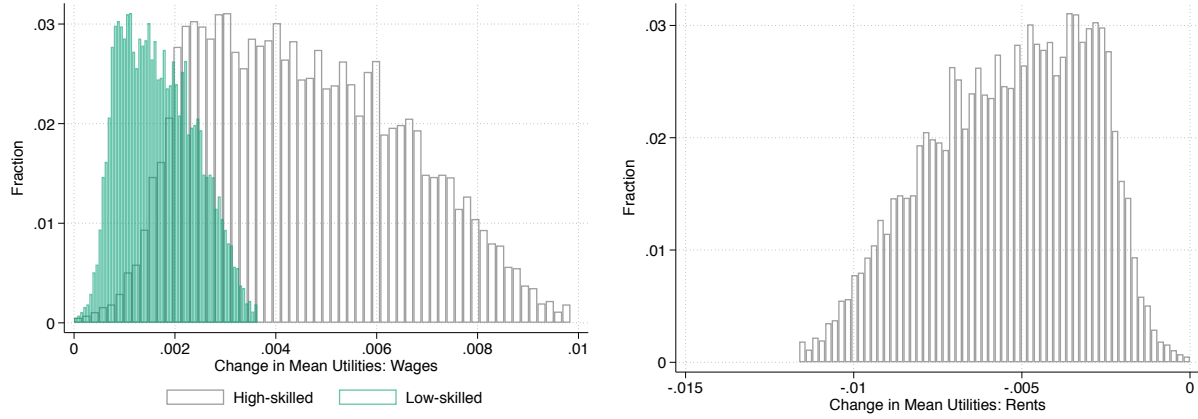
Notes: Treatment effects 5 to 8 years after firm entry announcement are in percent terms of the 0–10 minute ring around the entry site relative to the mean of the control group, which is the 20–30 minute ring around the entry site. Treatment effects are presented for four groups: low-skilled renters, low-skilled owners, high-skilled renters, and high-skilled owners. The lines to either side of the green circle and gray diamond for point estimates represent 95% confidence intervals. Control means of each group 5 to 8 years after firm entry announcement are in brackets next to the point estimates. The first outcome variable is whether an individual has never had bankruptcies. The second outcome variable is the share of incumbent residents who moved more than 1 mile away from their address when the firm entry is announced. The third outcome variable is the share of incumbent residents who moved more than 100 miles away, which proxies for moving out of the metropolitan area.

Figure B15: Quality of Incumbent's Current Neighborhood Outcomes

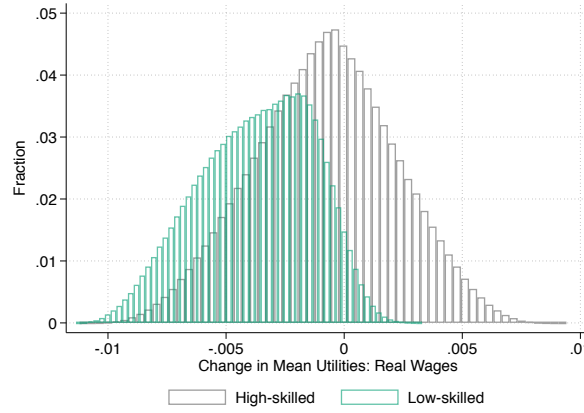


Notes: Treatment effects 5 to 8 years after firm entry announcement are in percent terms of the 0–10 minute ring around the entry site relative to the mean of the control group, which is the 20–30 minute ring around the entry site. Treatment effects are presented for four groups: low-skilled renters, low-skilled owners, high-skilled renters, and high-skilled owners. The lines to either side of the green circle and gray diamond for point estimates represent 95% confidence intervals. Control means of each group 5 to 8 years after firm entry announcement are in brackets next to the point estimates. The first and second outcome variable is the 1990 Income Per Capita and the 1990 Median Home Value of the Census Tract where individuals reside. The remaining outcome variables are upward mobility measures from Chetty et al. (2018) and based on the sample of children in the 1978–1983 birth cohorts who are U.S. citizen or authorized immigrants. Children's individual incomes are measured in 2014 and 2015, when they are between the ages of 31 and 37. The third and fourth outcome variable are the probability of reaching the top 1% of the national individual income distribution conditioning on the parental income at 50 and 75 percentile in the national household income distribution. The fifth outcome variable is the mean probability of reaching the top quintile of the national individual income distribution which is not conditioning on the parental income. The sixth, seventh and eighth outcome variables are the probability of reaching the top quintile of the national individual income distribution conditioning on the parental income at 25, 50 and 75 percentile in the national household income distribution.

Figure B16: Decomposition of Mean Utilities of Home-Work Location Choices for Renters



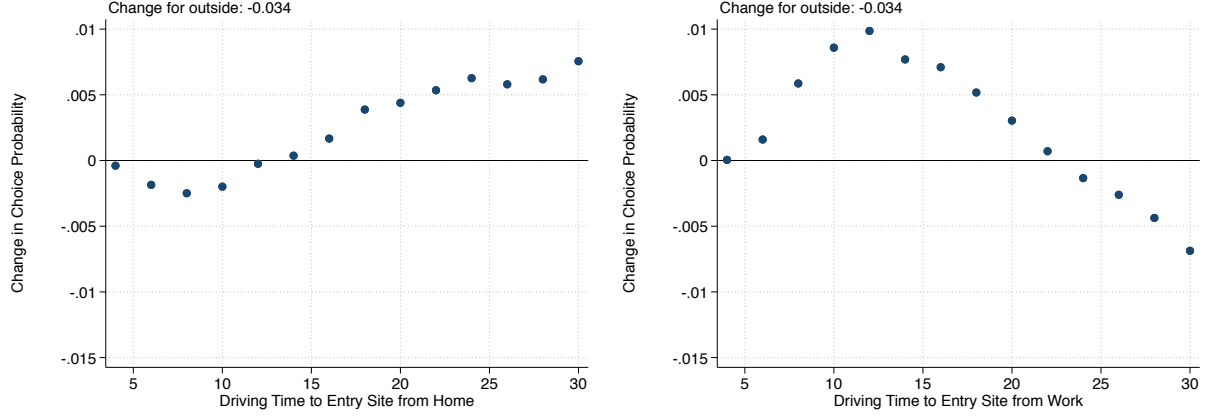
(a) Changes in mean utilities due to changes in wages (b) Changes in mean utilities due to changes in rents



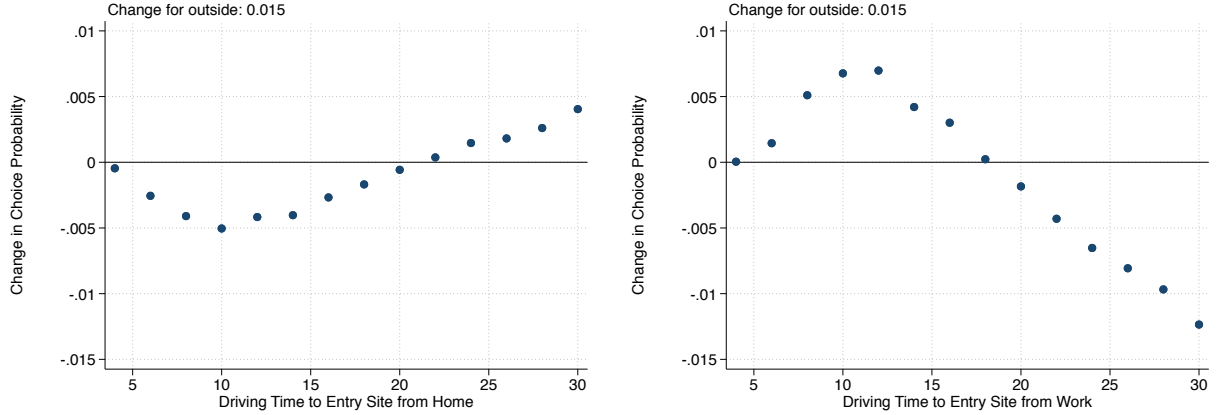
(c) Changes in mean utilities due to changes in real wages

Notes: This figure decomposes the changes in the estimated mean utility levels for home-work location choices inside the entry zones for renters,  $\Delta\delta_{i,j}^s$ , between period 1 prior to the entry and period 2 during which the firm entry happens. Panel (a) shows the distribution of the changes in mean utilities that are due to changes in wages in workplace neighborhoods,  $\Delta w_j^s$ , by skill group. Panel (b) shows the distribution of the changes in mean utilities that are due to changes in rents in residential neighborhoods,  $\Delta r_i$ . Panel (c) shows the distribution of the changes in mean utilities that are due to changes in real wages of each pair of home-work locations,  $\Delta w_j^s - \theta^R \Delta r_i$ , by skill group.

Figure B17: Changes in Choice Probabilities: High-skilled



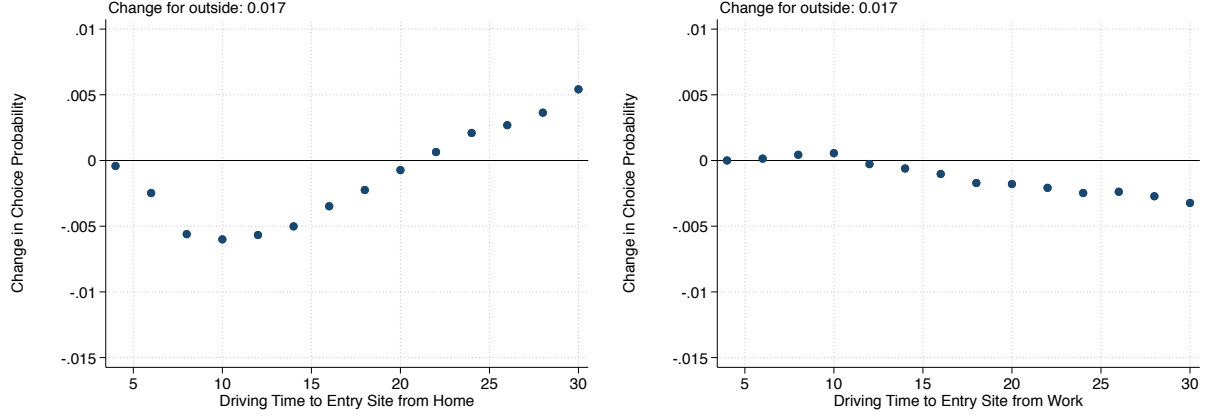
(a) High-skilled owners:  $\Delta p$  for residential neighborhoods (b) High-skilled owners:  $\Delta p$  for workplace neighborhoods



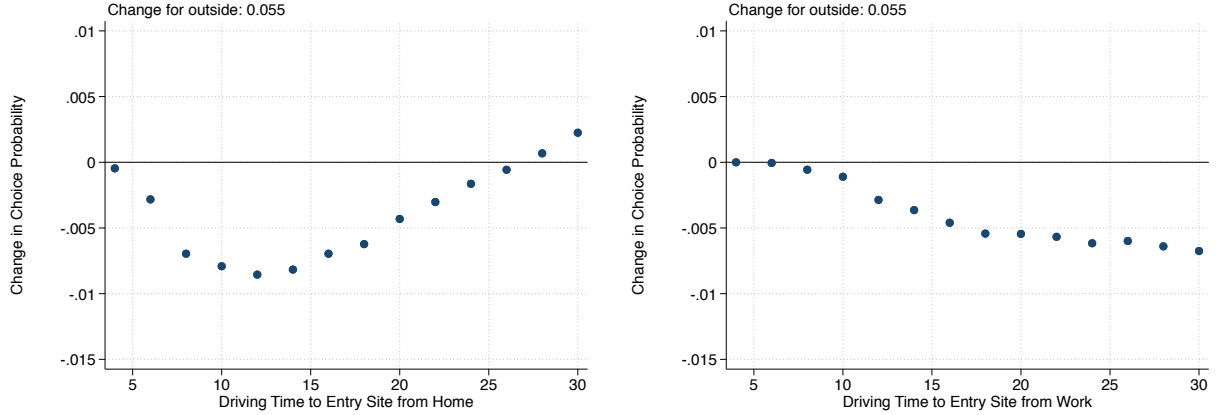
(c) High-skilled renters:  $\Delta p$  for residential neighborhoods (d) High-skilled renters:  $\Delta p$  for workplace neighborhoods

Notes: This figure plots the model predicted changes in high-skilled workers' choice probabilities between period 1 and 2 for residential and workplace neighborhoods by their driving time to the entry site. The x-axis is bins of driving time from residential or workplace neighborhoods to the entry site. We group neighborhoods into bins by every 2 mins in driving time to the entry site, and plot on the y-axis the total change in choice probabilities for residential or workplace neighborhoods in each bin of driving time across 129 firm entries in our estimation sample. The total change in the probability of choosing the outside option is written in the top left corner.

Figure B18: Changes in Choice Probabilities: Low-skilled



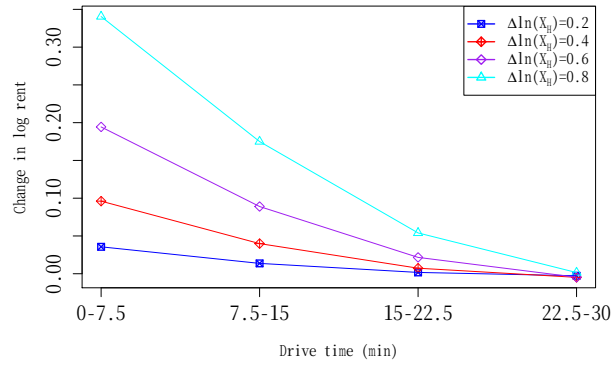
(a) Low-skilled owners:  $\Delta p$  for residential neighborhoods (b) Low-skilled owners:  $\Delta p$  for workplace neighborhoods



(c) Low-skilled renters:  $\Delta p$  for residential neighborhoods (d) Low-skilled renters:  $\Delta p$  for workplace neighborhoods

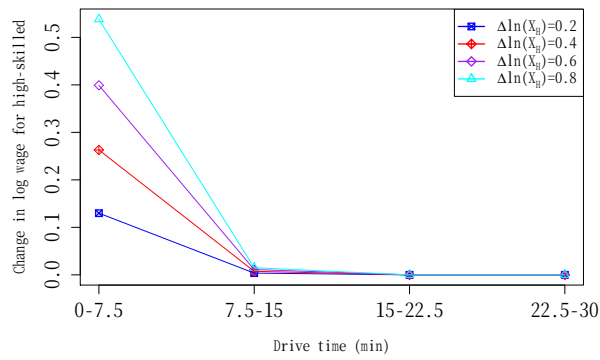
Notes: This figure plots the model predicted changes in low-skilled workers' choice probabilities between period 1 and 2 for residential and workplace neighborhoods by their driving time to the entry site. The x-axis is bins of driving time from residential or workplace neighborhoods to the entry site. We group neighborhoods into bins by every 2 mins in driving time to the entry site, and plot on the y-axis the total change in choice probabilities for residential or workplace neighborhoods in each bin of driving time across 129 firm entries in our estimation sample. The total change in the probability of choosing the outside option is written in the top left corner.

Figure B19: Changes in Rents

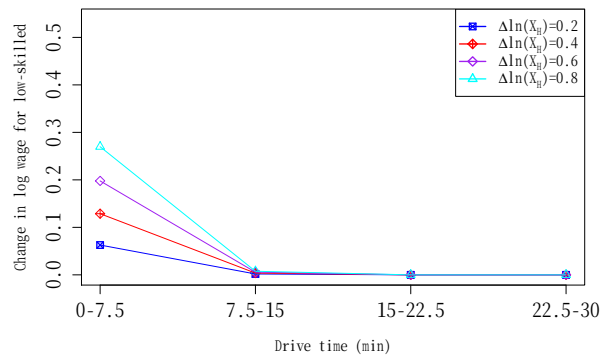


Notes:

Figure B20: Changes in Wages



(a) Change in wages for high-skilled workers

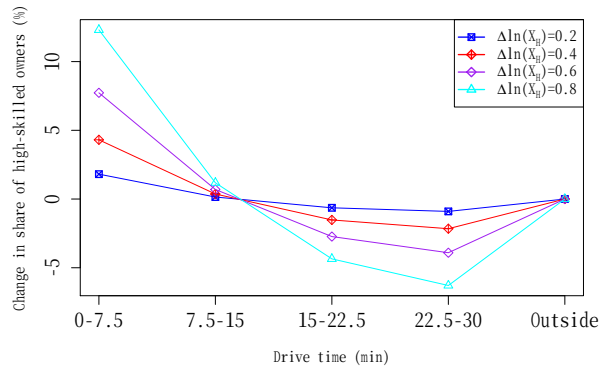


(b) Change in wages for low-skilled workers

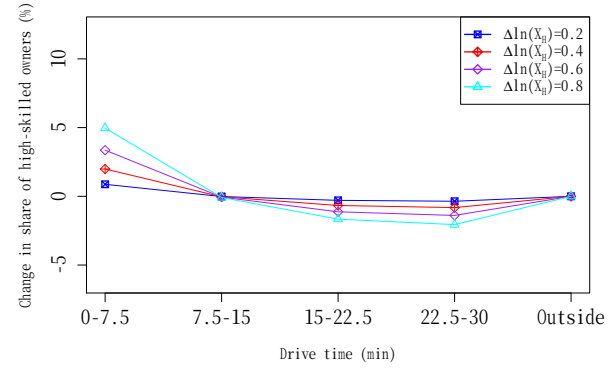
Notes:



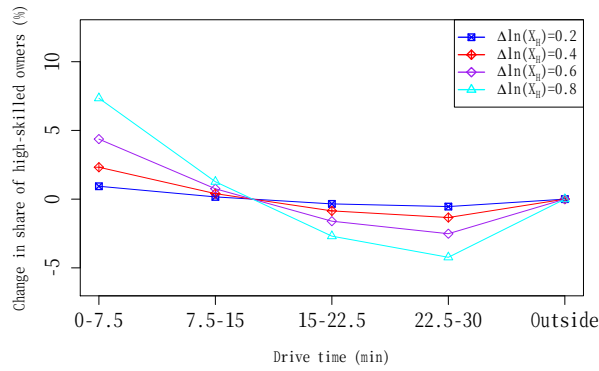
Figure B21: Residential Choice of High-skilled Owners



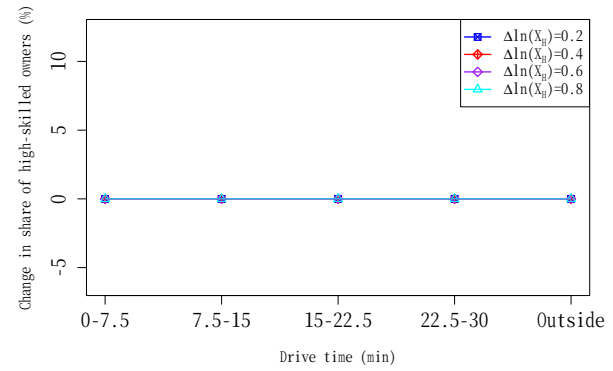
(a) High-skilled owners



(b) High-skilled owners born within 0-15 mins from entry site



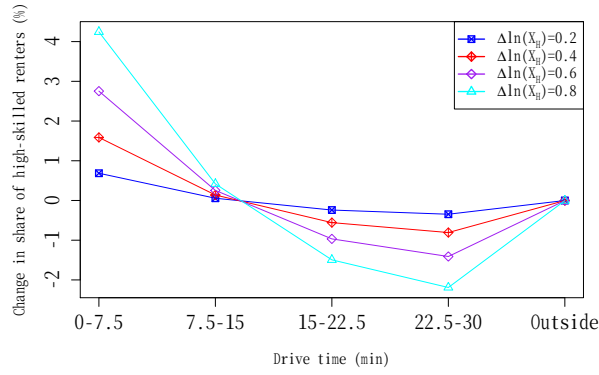
(c) High-skilled owners born within 15-30 mins from entry site



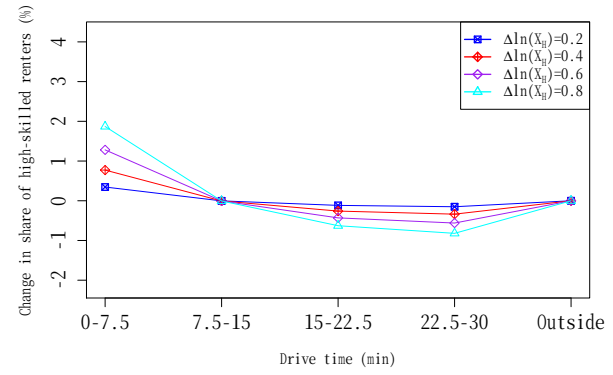
(d) High-skilled owners born outside

Notes:

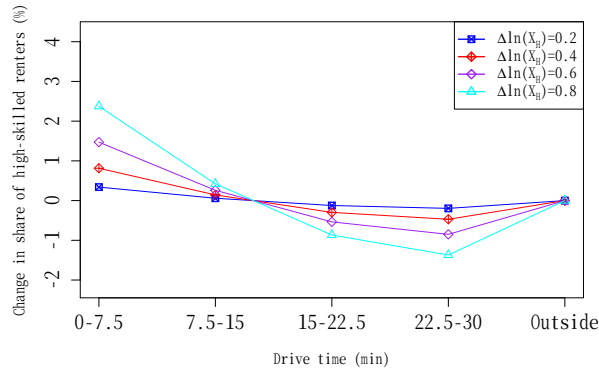
Figure B22: Residential Choice of High-skilled Renters



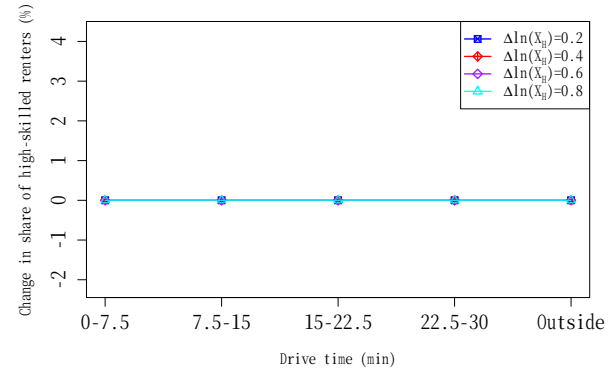
(a) High-skilled renters



(b) High-skilled renters born within 0-15 mins from entry site



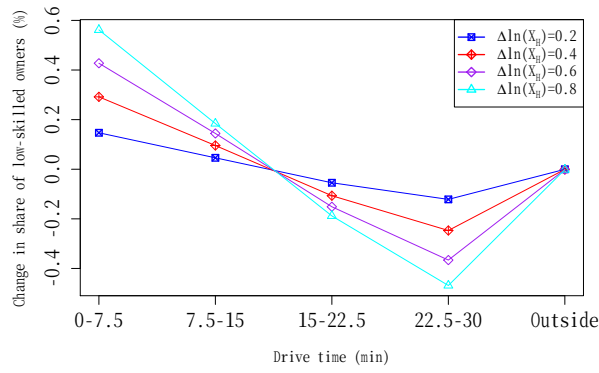
(c) High-skilled renters born within 15-30 mins from entry site



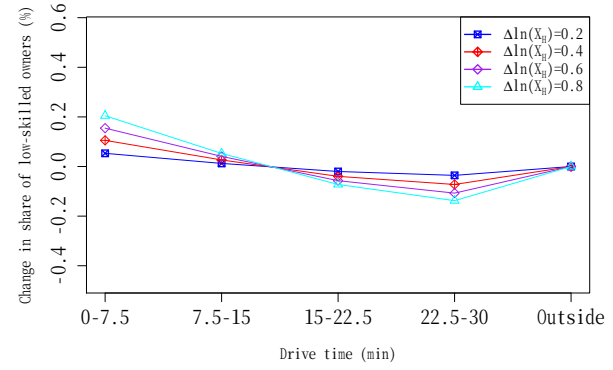
(d) High-skilled renters born outside

Notes:

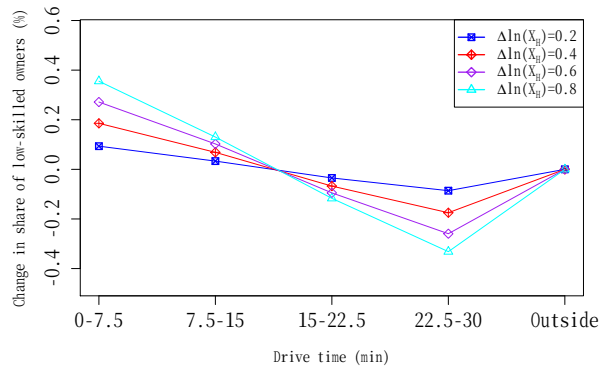
Figure B23: Residential Choice of Low-skilled Owners



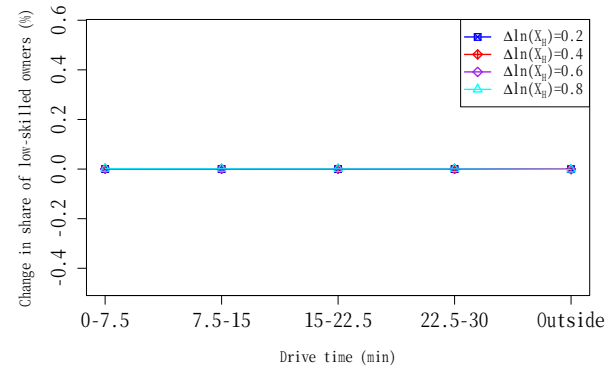
(a) Low-skilled owners



(b) Low-skilled owners born within 0-15 mins from entry site



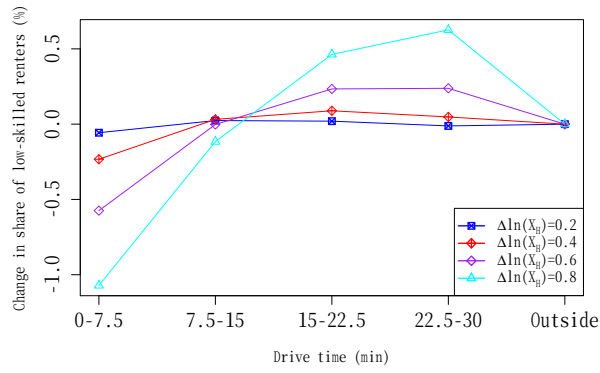
(c) Low-skilled owners born within 15-30 mins from entry site



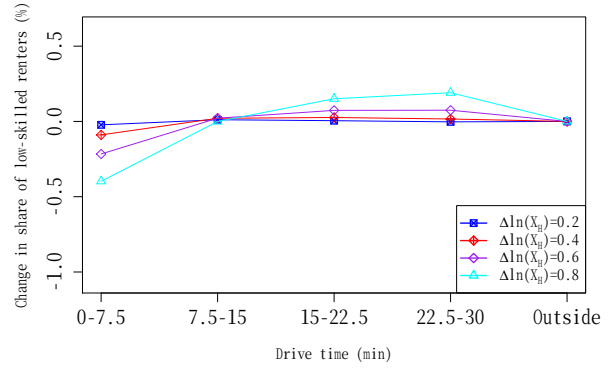
(d) Low-skilled owners born outside

Notes:

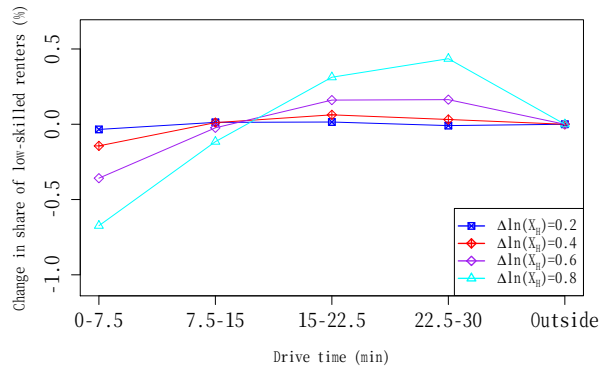
Figure B24: Residential Choice of Low-skilled Renters



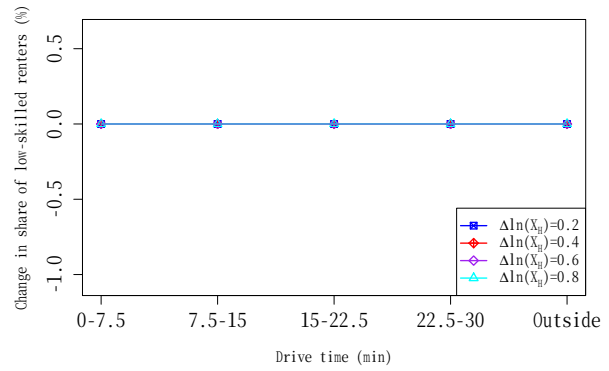
(a) Low-skilled renters



(b) Low-skilled renters born within 0-15 mins from entry site



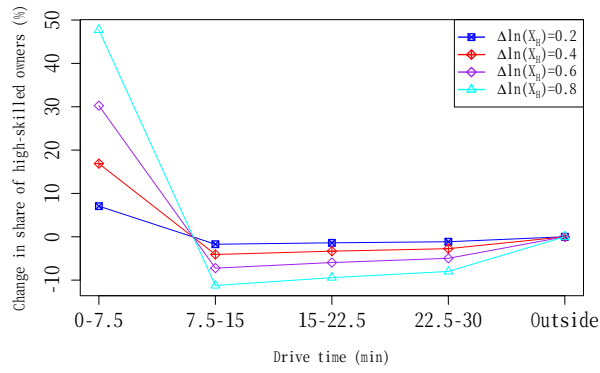
(c) Low-skilled renters born within 15-30 mins from entry site



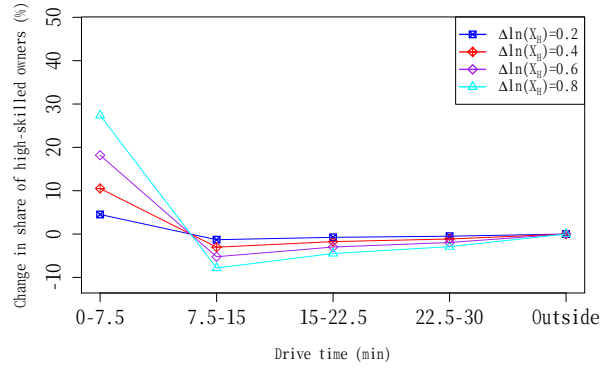
(d) Low-skilled renters born outside

Notes:

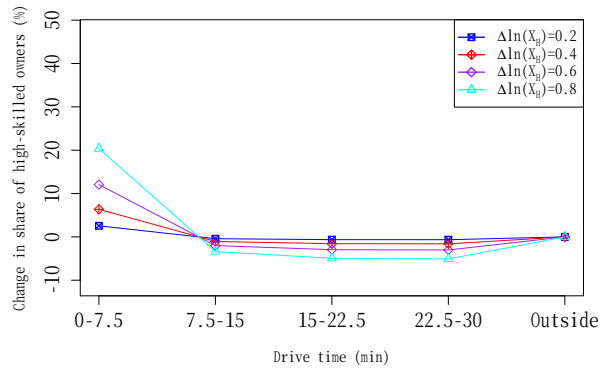
Figure B25: Workplace Choice of High-skilled Owners



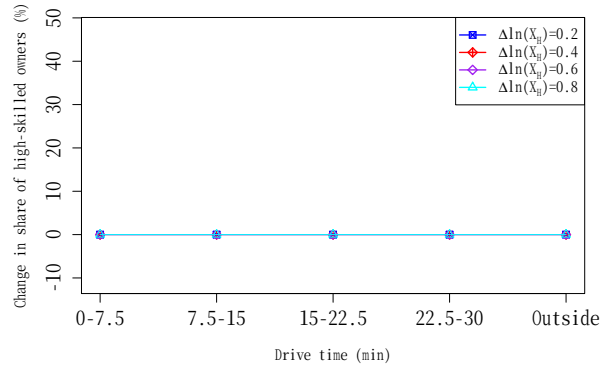
(a) High-skilled owners



(b) High-skilled owners born within 0-15 mins from entry site



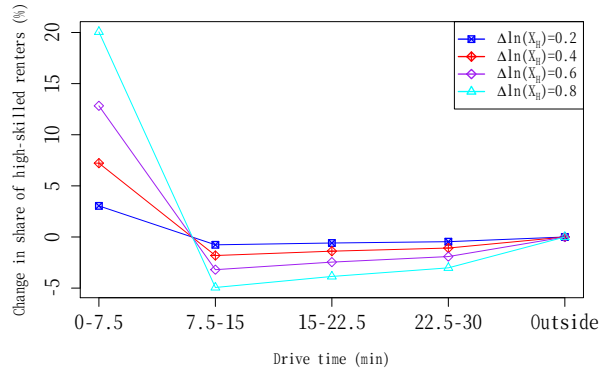
(c) High-skilled owners born within 15-30 mins from entry site



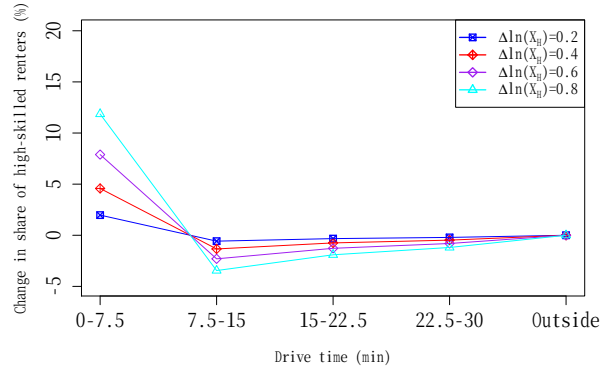
(d) High-skilled owners born outside

Notes:

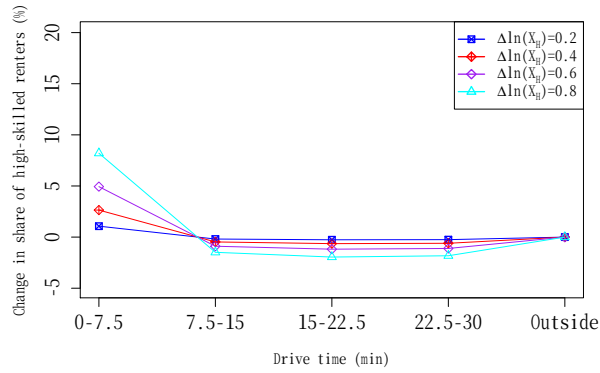
Figure B26: Workplace Choice of High-skilled Renters



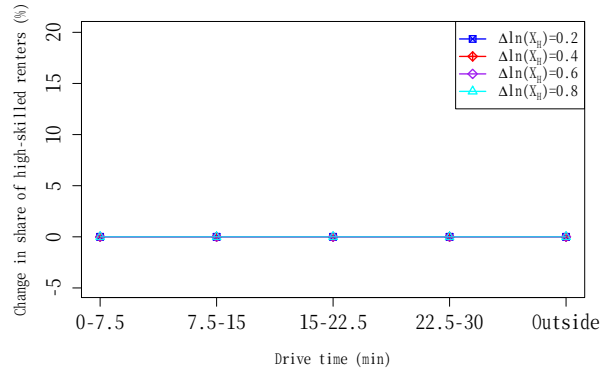
(a) High-skilled renters



(b) High-skilled renters born within 0-15 mins from entry site



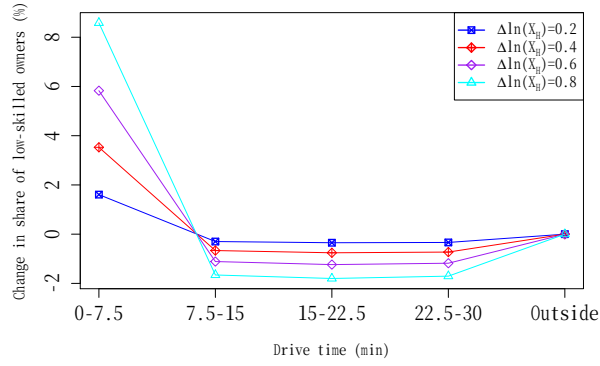
(c) High-skilled renters born within 15-30 mins from entry site



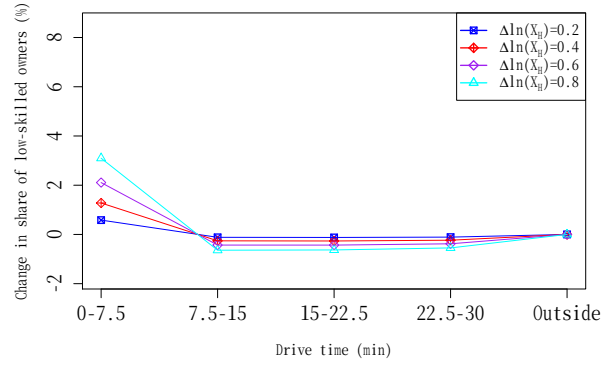
(d) High-skilled renters born outside

Notes:

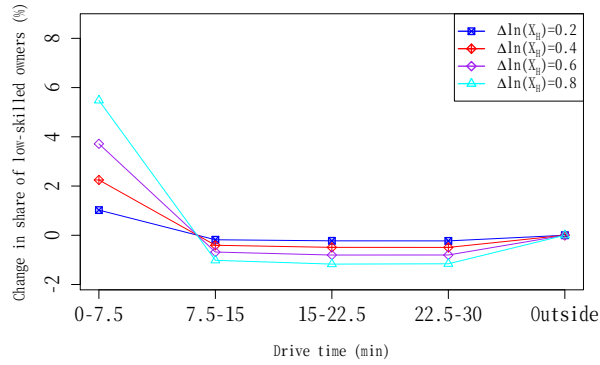
Figure B27: Workplace Choice of Low-skilled Owners



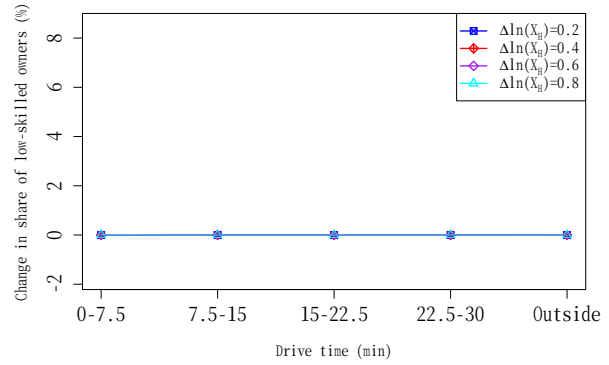
(a) Low-skilled owners



(b) Low-skilled owners born within 0-15 mins from entry site



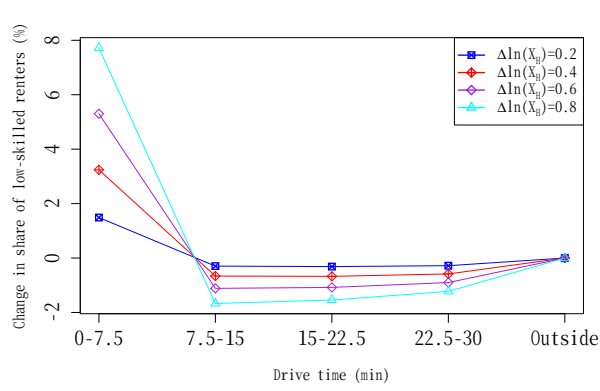
(c) Low-skilled owners born within 15-30 mins from entry site



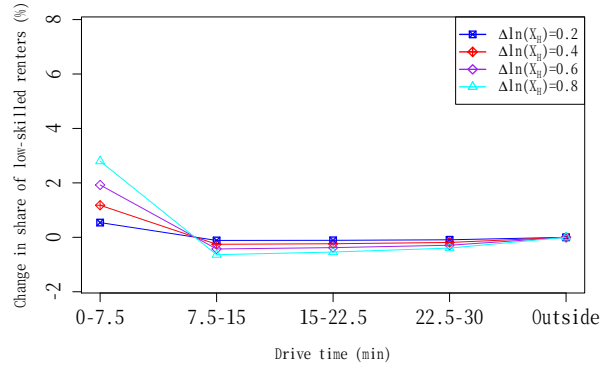
(d) Low-skilled owners born outside

Notes:

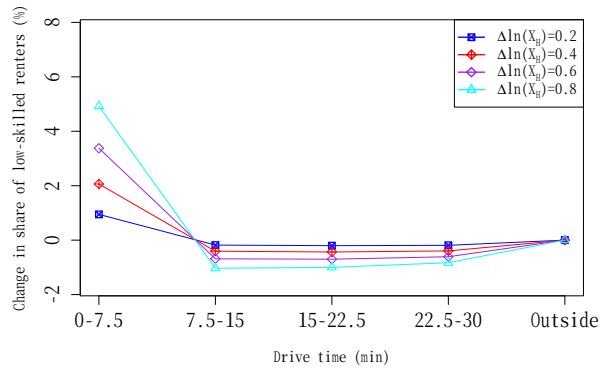
Figure B28: Workplace Choice of Low-skilled Renters



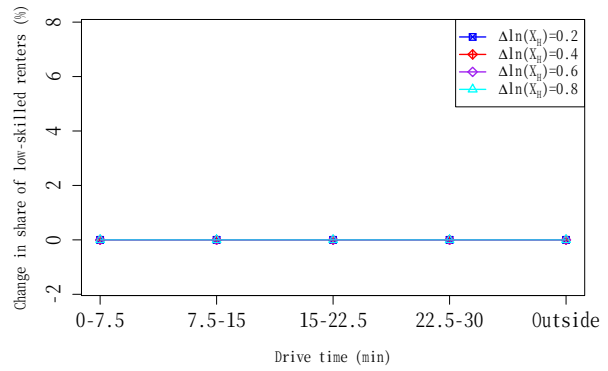
(a) Low-skilled renters



(b) Low-skilled renters born within 0-15 mins from entry site



(c) Low-skilled renters born within 15-30 mins from entry site

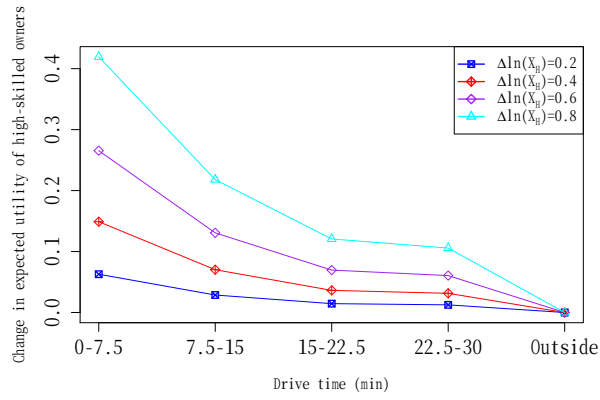


(d) Low-skilled renters born outside

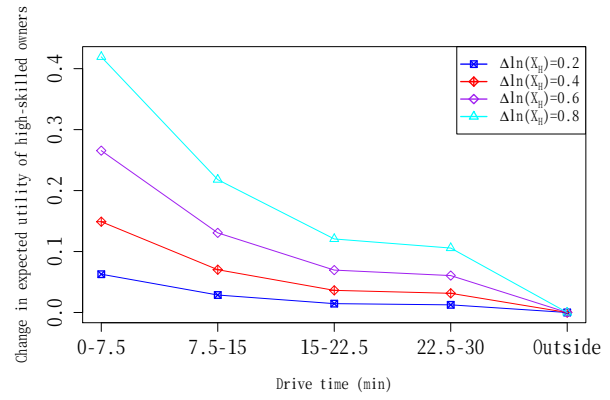
Notes:



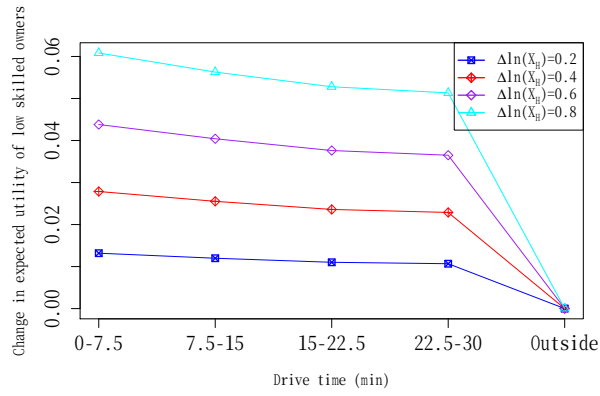
Figure B29: Change in Expected Utility



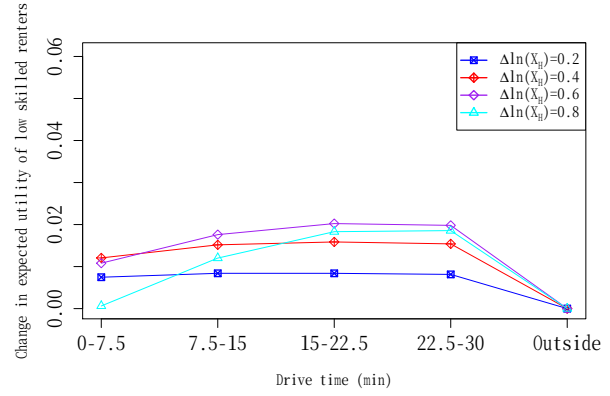
(a) High-skilled owners



(b) High-skilled renters



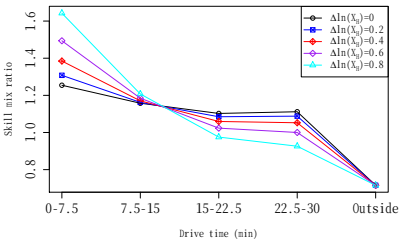
(c) Low-skilled owners



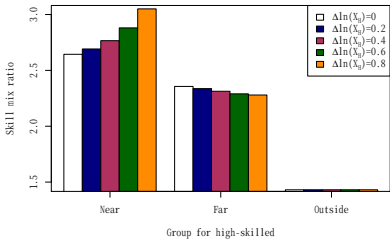
(d) Low-skilled renters

Notes:

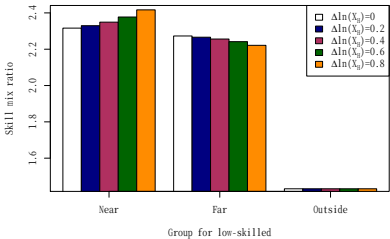
Figure B30: Change in Neighborhood Skill-Mix



(a) Residential neighborhood skill-mix vs driving time



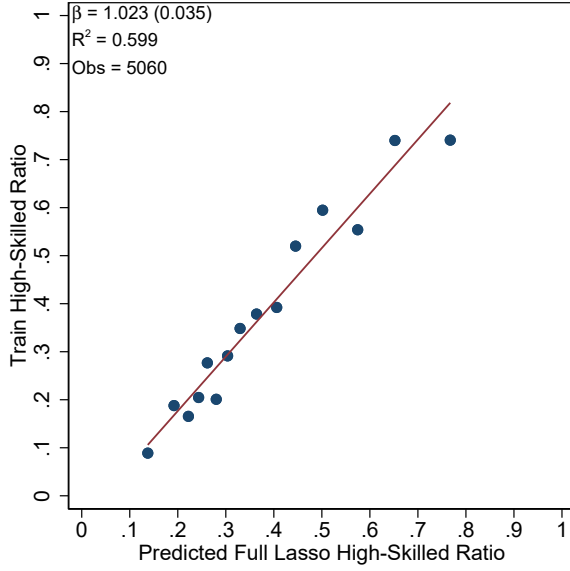
(b) Avg. residential neighborhood skill-mix of high-skilled workers



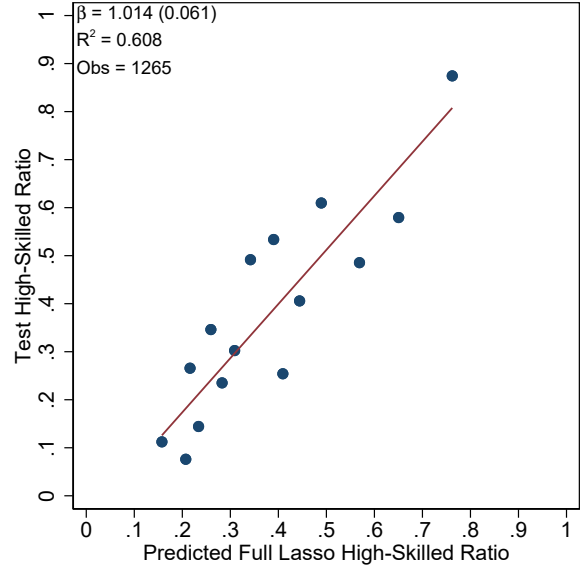
(c) Avg. residential neighborhood skill-mix of low-skilled workers

Notes:

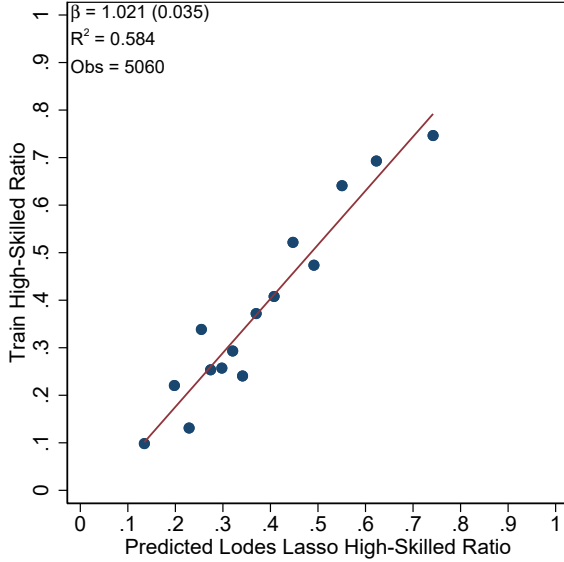
Figure B31: Validation of Commute Pattern Model



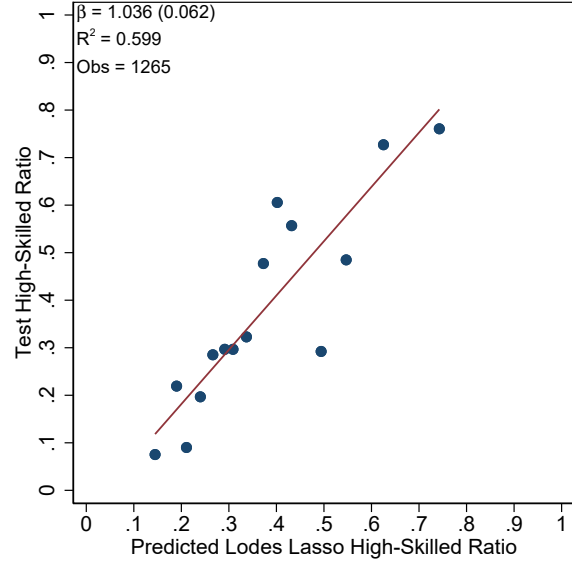
(a) Training: Lasso with full set of predictors



(b) Test: Lasso with full set of predictors



(c) Training: Lasso with LODES predictors



(d) Test: Lasso with LODES predictors

Notes: This figure presents bin-scatter plots of actual high-skilled ratios  $\mathcal{S}_{i,j}^H$  for pairs of home-work Tracts  $(i, j)$  against the model predicted high-skilled ratios  $\hat{\mathcal{S}}_{i,j}^H$  in our training and test samples, respectively. The X-axis is the predicted high-skilled ratio, and the Y-axis is the actual high-skilled ratio. We divide the X-variable into 15 bins and plot the average value of the Y-variable for each bin. We report the slope of regression, along with robust standard error in parentheses, and  $R^2$  in the top left corner. Regressions are weighted by the aggregated personal weights in each pair of Tracts. In Panel (a), the sample is our training sample which makes up 80% of all home-work Tract pairs from 2009 NHTS. The model is a Lasso Logistic model using the full set of predictors. In Panel (b), the sample is our test sample which makes up 20% of all home-work Tract pairs from 2009 NHTS. The model is a Lasso Logistic model using the full set of predictors. In Panel (c), the sample is our training sample. The model is a Lasso Logistic model using the LODES predictors. In Panel (d), the sample is our test sample. The model is a Lasso Logistic model using the LODES predictors.

## C Data Construction

### C.1 Homeownership Imputation

To impute home ownership, we first match addresses from Infutor with addresses from tax assessor’s records of property characteristics and housing transaction data from CoreLogic. Details of this merge are provided in Appendix C.1.1. Next, we construct an address-year panel with property owner names of each address in each year using both tax records and transactions data from CoreLogic. More details for creating this owner panel are provided in Appendix C.1.2. Finally, we compute the Jaro-Winkler string distance between the last names of individuals from Infutor and the last names of property owners from CoreLogic in each year. An individual is classified as a homeowner in a given year if the string distance between his last name and any last name of the owners at his address in that year is greater or equal to 0.9. See Appendix Section C.1.3 for robustness to other cutoffs. Some of the data construction was performed jointly with Blattner and Nelson (2020) and we have jointly written this data appendix on imputing homeownership for Infutor panelists. Some cleaning of the raw Infutor migration history data is joint with Diamond et al. (2020). See their Appendix Section A.2 for additional details.

Figure B1 shows the time series comparison between the homeownership rate from Censuses (retrieved from FRED) and our imputed homeownership rate from the Infutor-CoreLogic linked data. We consistently under predict homeownership, but over time, the gap between our homeownership rate and the Census homeownership rate becomes smaller. Figure B2 shows we can capture over 95% of the variation in homeownership rate at the Census Tract level in different years.

#### C.1.1 Infutor-CoreLogic Address Merge

We first convert all distinct addresses in both Infutor and CoreLogic into clean, standardized versions. Note that addresses in CoreLogic are uniquely identified by the APN, county FIPS, and APN sequence number. We create 5-digit zip codes and 2-digit state codes, we remove all special characters and spaces from street names, split apartment numbers into numeric and non-numeric parts, and standardize city names using common abbreviations suggested by USPS. We also create a second city variable based on the preferred spelling of a city name by USPS (preferred city). We then keep all unique addresses in Infutor and CoreLogic respectively. We then merge the two address data sets using the following iterative procedure (at each step we remove uniquely matched observations).

- All address variables (street name and number, apartment number, zip code, city, state);
- All variables but only numeric portion of apartment;
- All variables using preferred city name;
- All variables using preferred city name but only numeric portion of apartment;

- All variables leaving out zip code;
- All variables leaving out zip and using preferred city names;
- All variables leaving out zip but only numeric portion of apartment;
- All variables but dropping apartment number if and only if there is no apartment number in CoreLogic. This step captures cases where the CoreLogic address is the whole apartment complex but Infutor shows addresses for individual apartments.
- Same as previous but using preferred city.

A note on apartment numbers. We observe cases where multi-unit buildings have a single owner in CoreLogic, and hence a missing apartment number, but we observe multiple renters of the individual units in Infutor. If we forced a match on apartment number, we would not match the multi-unit buildings since apartment number is missing in one dataset but not the other. This would imply that we would potentially miss out on classifying a large number of renters in Infutor. For this reason, we allow a final merge round that does not condition on apartment number. However, this implies that we generate a number of non-unique matches since now one CoreLogic address is associated with multiple Infutor addresses. This non-uniqueness however does not pose a problem for our owner-renter imputation (see details below).

Table A1 provides the fraction of Infutor addresses that we successfully match to CoreLogic by state. We get low match rates (as % of Infutor addresses matched) for a handful of states. We confirmed with CoreLogic customer relations representatives that this is driven by low CoreLogic coverage for these states as the deed records in these states have systems in place that make the digitization of the records more difficult.

### C.1.2 Construction of Owner Panel

To construct the ownership variable, we first use CoreLogic transaction and tax records to establish the owners for each address-year combination. We find that historical tax records to be very incomplete, so we rely on 2017 tax records and reconstruct ownership based on past transactions. We proceed in the following steps:

- Create a bookend based on the most recent transaction and tax entry. Note that for transactions, we use the buyer in the last transaction in that year. That is, our imputed owner variable will be the owner “as of the end of the year”. If there is a tax and transaction entry in the same year, the transaction data takes precedence.
- Fill in forward: If the most recent bookend is in a year prior to 2017 (the end of our CoreLogic sample), we assume that owner remains the same until 2017.
- Fill in backward: First, we create additional data points from prior transactions. Every time there is a transaction, we say that the last buyer in that year is the owner for the year. We

then fill in data between the transactions: If we have a transaction in year  $t$ , then the owner at  $t - 1$  is the seller of the first transaction at  $t$ . For  $t - 2$ ,  $t - 3$  etc. we then assume the owner from  $t - 1$  remains the owner (until we hit the next transaction). For tax-only entries, we just assume that the owner at  $t$  remains the owner.

For robustness, we use an alternative fill-in procedure that only uses transaction data:

- We find the last buyer in the earliest transaction for that address. That’s our first owner.
- That person remains the owner until there is another transaction. Then the last buyer in that year becomes the new owner and remains the owner until there is another transaction.

When matching against names in Infuturs, we match against owner names from both fill-in procedures if they conflicting information. We repeat this procedure for the property type.

In the final step, we compare whether the name in Infutor matches the owner name on the deed record. We consider both first and second owners. We first clean names in both data sets and remove any special characters and spaces. We then compute the Jaro-Winkler string distance between the last names of individuals from Infutor and the last names of property owners from CoreLogic in each year. An individual is classified as a homeowner in a given year if the string distance between her last name and any last name of the owners at his address in that year is greater or equal to 0.9.

### C.1.3 Robustness to Distance Cut-off

We also explore using string distance cutoff values of 0.8, 0.85 and 0.95. We check a random sample of 300 cases that are inconsistent for the different cutoffs. The majority of cases that are a match for 0.8 but not a match for 0.85 are not true matches. Cases that are a match for 0.85, but not for 0.9 are noisy, with about 50% of them being true matches. Cases that are a match for 0.9, but not for 0.95 are mostly true matches. Only 0.88% of all cases have inconsistencies using different cutoff values of 0.85, 0.9, and 0.95. In addition, when an Infutor last name has more than 3 characters and is fully contained in CoreLogic owner last name and vice versa, we adjust the string distance to be 1 and call it a match. This deals with cases where an individual’s last name is contained in a corporation’s name, such as a trust or an estate. This also deals with the structure of some Hispanic names, for example, “Ocasio Cortez” & “Cortez”, “Jimenez de Armendariz” & “Armendariz”, but would not match “de” with “de Colon”.

## C.2 Education/Skill Imputation

To impute education, first we use demographic and housing characteristics from the ACS to build a model, then we predict out of sample using the same set of variables we observe in our Infutor-CoreLogic linked data, and finally we Bayesian update based on an individual’s Census block group. We classify individuals as high-skilled or low-skilled based on their education level: those with at least a college degree are high-skilled and those without are low-skilled. While education and skill

are not identical, education may be viewed as a proxy for skill, and we will use the two terms interchangeably in this paper.

First, we first build a model that predicts whether a person has a college degree or above using the 1% micro data of individuals aged 25 to 65 from the 2010 ACS. After splitting the data into a 90% training sample, and 10% test sample, we estimate a flexible logistic model of education as a function of demographic and housing characteristics. Demographics used are gender, marital status, race, immigrant status, and age. Housing characteristics include homeownership status, number of bedrooms, number of rooms, bedroom-to-room ratio, property type, building age, and house value (if a homeowner) or monthly rent (if a renter). We allow for two-way interactions between these variables and add state fixed-effects and their interaction with these variables. Column 3 of Table A2 shows the estimated coefficients in our main model. Using the 10% test sample for validation, we find the model does quite well and achieves a 73.8% predictive accuracy<sup>41</sup>.

Next, we use the skill prediction model to predict out-of-sample on the Infutor panel of individuals based on their observed characteristics. Housing characteristics come from merging Infutor with CoreLogic, and details of this merge are provided in Appendix C.1.1. We obtain the same set of variables used in our skill prediction model from the Infutor-CoreLogic merged data set, and then apply the coefficients estimated in our model to obtain the baseline probability of being high-skilled for each individual<sup>42</sup>.

Finally, we Bayesian update each individual’s baseline high-skilled probability with education data from their 2010 Census block group, similar to Elliott et al. (2009), Consumer Financial Protection Bureau (2014), and Diamond et al. (2019)<sup>43</sup>. Bayes’ Rule gives the probability that an individual with demographic and housing characteristics  $d$  residing in geographic area  $g$  belongs to skill group  $s$ :

$$Pr(s|g, d) = \frac{Pr(s|d)Pr(g|s, m)}{\sum_{s' \in \{H, L\}} Pr(s'|d)Pr(g|s', m)},$$

where  $Pr(s|d)$  is the baseline probability of belonging to skill group  $s \in \{H, L\}$  given an individual’s demographic and housing characteristics  $d$ , and  $Pr(g|s, m)$  is the proportion of the population of gender  $m$  belonging to skill group  $s$  who live in Census block group  $g$  obtained from the 5-Year Summary File of the 2008–2012 ACS<sup>44</sup>. A simple proof of the Bayes’ Rule equation is in Appendix

<sup>41</sup>A person is classified as high-skilled if he has a probability of having a college degree or above that is equal or greater than 50%, otherwise he is classified as low-skilled.

<sup>42</sup>There are some missing values in our linked Infutor-Corelogic data, so to maximizing the sample size, we use a dummy variable trick. We create a dummy for each independent variable that equals 1 if the value of that variable is missing. To keep the set of variables (including missing value dummies) consistent between ACS and our linked Infutor-Corelogic data, we randomly drop 0.1% observations from variables in the ACS and generate missing value dummies in the ACS sample. To fill in missing values of each variable, we use the median of the rest of the non-missing observations from ACS, and assign the median to the observations with missing value in the corresponding variable from our linked Infutor-Corelogic data.

<sup>43</sup>Diamond et al. (2019) demonstrates that combining geography- and name-based information into a single proxy probability for race/ethnicity significantly outperforms traditional classification methods based on names or geography alone.

<sup>44</sup>In practice Census block group level information on gender by skill group is available for 94.5% of our sample. For

C.2.1. An assumption necessary for the validity of the Bayesian updating procedure is that the probability of living in a given geographic area, given one’s gender and level of skill, is independent of all other demographic and housing characteristics<sup>45</sup>.

### C.2.1 Proof for Bayesian Updating Formula for Skill Level

The assumption that the probability of living in a given geographic area, given one’s gender and level of skill, is independent of one’s demographic and housing characteristics means that:

$$\Pr(g|s, d, m) = \Pr(g|s, m).$$

Also note  $m \in d$  since gender is part of an individual’s characteristics used to predict his baseline probability of skill level. Hence,

$$\Pr(s|m, d) = \Pr(s|d).$$

*Proof.*

$$\begin{aligned} \Pr(s|g, d, m) &= \frac{\Pr(s, g, d, m)}{\sum_{s' \in \{L, H\}} \Pr(s', g, d, m)} \\ &= \frac{\Pr(g|s, d, m) \Pr(s|m, d) \Pr(m, d)}{\sum_{s' \in \{L, H\}} \Pr(g|s', d, m) \Pr(s'|m, d) \Pr(m, d)} \\ &= \frac{\Pr(g|s, m) \Pr(s|m, d)}{\sum_{s' \in \{L, H\}} \Pr(g|s', m) \Pr(s'|m, d)} \\ &= \frac{\Pr(g|s, m) \Pr(s|d)}{\sum_{s' \in \{L, H\}} \Pr(g|s', m) \Pr(s'|d)} \end{aligned}$$

□

## C.3 Census Tract Level House Price Indices

We construct a hedonic house price index at the Census Tract level using CoreLogic transaction data from 1990–2017. We restrict our sample to arm’s-length transactions of single-family homes and condominiums. Sale prices for houses are normalized to 2014 real dollars. Our method directly controls for a suite of housing characteristics from the tax assessor’s property records provided by CoreLogic. For each census tract  $i$ , we construct  $\text{HPI}_{i,t}$  using the following regression specification:

---

the rest of the sample, we use information at the Census tract (4.1%) and 5-digit ZIP code levels (1.1%), whichever one is first available in the order listed. We set the posterior probabilities equal to the baseline probabilities for the remaining 0.3% of the sample.

<sup>45</sup>This assumption implies that high-skilled males with a certain set of demographic characteristics are just as likely to live in a certain neighborhood as high-skilled males in general.



$$\ln(\text{Price}_{h,i,t}) = \text{HPI}_{i,t} \text{Year}_t + \mathbf{X}_{h,t} \beta + \varepsilon_{h,i,t}, \quad (25)$$

We regress the log of transaction price of home  $h$  in Tract  $i$  in year  $t$ ,  $\ln(\text{Price}_{h,i,t})$ , on calendar year dummies,  $\text{Year}_t$ , and  $\mathbf{X}_{h,t}$ , which are a set of property characteristics which include whether the property is a condominium, number of bathrooms, bedrooms to bathrooms ratio, living area size, building age, and age<sup>2</sup>. The estimates  $\text{HPI}_{i,t}$  provides the house price index in Census Tract  $i$  in calendar year  $t$ . This gives us a Tract by year panel of house price indices.

## C.4 Driving Time

We quantify the distance from each neighborhood to the firm entry location using historical driving time, which is obtained by estimating a speed model using the 1995 National Household Travel Survey (NHTS)<sup>46</sup>, similar to [Couture \(2016\)](#) and [Su \(2019\)](#). Then we use the model parameters to predict 1990 travel speed using 1990 Census characteristics.

First, we estimate the following speed model:

$$\log(\text{Speed}_{i,j,t}) = \beta_{0,t} + \beta_{1,t} \log(\text{Distance}_{i,j}) + \beta_{2,t} \log(\text{Distance}_{i,j})^2 + \mathbf{X}_{i,j,t} \beta_{3,t} + \varepsilon_{i,j,t}, \quad (26)$$

where  $i$  is the origin Census tract where the trip begins,  $j$  is the destination Census tract when the trip ends, and  $t$  is the year in which the trip is taken. The terms  $\log(\text{Distance}_{i,j})$  and  $\log(\text{Distance}_{i,j})^2$  represent the fact that speed is a function of distance, since people generally tend to travel faster on longer trips.  $\mathbf{X}_{i,j,t}$  are neighborhood characteristics that may affect speed, including population density, employment density, and log median household income<sup>47</sup>. This model is estimated using data for morning and evening commute trips from the 1995 NHTS<sup>48</sup>. We apply sampling weights for trips to make sure trips in our sample are nationally representative. We only use an 80% sample of NHTS trips to estimate the model, with the remaining 20% of the data held out as a test sample to evaluate model fit.

Next, we apply the speed model parameters to 1990 Census tract characteristics to impute 1990 speed:

$$\widehat{\log(\text{Speed}_{i,j,1990})} = \hat{\beta}_{0,1995} + \hat{\beta}_{1,1995} \log(\text{Distance}_{i,j}) + \hat{\beta}_{2,1995} \log(\text{Distance}_{i,j})^2 + \mathbf{X}_{i,j,1990} \hat{\beta}_{3,1995}. \quad (27)$$

In [Figure B3](#), we evaluate the accuracy of our speed model in two ways using the 20% test sample, and the 2019 contemporaneous travel time from Google Maps. The model predictions closely match the empirical data. [Figure B3a](#) plots the predicted travel speed against the actual

<sup>46</sup>We use 1995 NHTS instead of 1990 NHTS because the former has richer neighborhood demographics.

<sup>47</sup>These are the variables with the highest univariate correlation with travel speed. We only have these variables for the origin location and not the destination location due to data limitations. Higher population density and employment density are associated with more congestion and lower speed. The parameters of the model vary by year as because traffic conditions may evolve over time due to changes including economic development.

<sup>48</sup>To identify home to work commutes, we use trips that depart between 6:30am and 10:30am. To identify work to home commutes, we use trips that depart between 4:30pm and 8:30pm.

travel speed from the 20% test sample. The slope of the regression line is 0.972 and the  $R^2$  is 0.66, which means that our model predicts the actual speed quite well. In Figure B3b, using the Google Maps Distance Matrix API, we compute the amount of time it takes to drive from the firm entry location to the centroid of each Tract within a 25-mile radius around the entry site without traffic at 5 A.M. in July 2019. By shutting down traffic in the early morning, this measure is a proxy for commuting cost from neighborhoods to firm’s location due to traffic infrastructure, and not due to traffic conditions which could respond to the firm entry. We then plot the predicted 1990 travel time against the Google Maps travel time. The  $R^2$  of the regression line is 0.85, which suggests our predicted historical travel time captures the variation in commuting cost well. These two methods validate that our speed model yields accurate predictions.

Figure B4 shows the distribution of driving time from the firm entry location to the centroid of each Tract within 40 minutes of driving time. In our main analysis, we divide Tracts around the firm entry site into three rings. The inner ring of Tracts 0–10 minutes away from the firm entry is the most treated group, the 10–20 minute ring is the second most treated group, and the outer ring 20–30 minutes away from the entry is the control group. Figure B4 shows that there are a sufficient number of Tracts within 30 minutes of the firm entry sites. The distribution of Tracts is smooth across the range of driving time considered, so there does not exist discontinuities around 30 minutes. There are also a sufficiently large number of neighborhoods in each ring so that we have enough power to measure outcomes accurately.

## C.5 Data Implementation for Estimation

### C.5.1 Choice Probabilities by Skill

LODES Origin-Destination (OD) files provide us with bilateral commuting flows between residential and workplace Census Tracts. Since the data does not include worker education level, we impute commuting flows by skill group using a Lasso Logistic model to predict the share of high-skilled workers commuting between a pair of residential and workplace neighborhoods based on their neighborhood characteristics, obtained from LODES, Censuses/ACS and ZBP. Our training sample comes from a nationally representative sample of commuters with known education levels from the 2009 NHTS. We then use the model to predict the share of high-skilled workers commuting between each pair of Census Tracts  $(i, j)$  within 50 miles from each firm entry site. We crosswalk the commuting flows from Tract to ZIP Code level. In period 1 and 2, we observe the total population by skill who choose to live and work inside the 50-mile radius around each firm’s entry site. We also observe the number of workers who choose each pair of home-work ZIP Codes by skill group. Hence, we can calculate the choice probability,  $\pi_{i,j,t}^s$ , for each pair of home-work ZIP Codes by skill group in each period. More details on how impute commute pattern by skill group are provided below.

### Imputation of Commute Patterns by Skill Group

We predict  $n_{i,j}^s$ , the number of workers in skill group  $s$  who live in Census Tract  $i$  and work in Census Tract  $j$  within 50 miles of each firm’s entry site.

## Data

1. **Commuting Patterns** — LEHD Origin-Destination Employment Statistics (LODES) provides Census Block-to-Block worker commuting flows from 2002 – 2015, including Origin-Destination (OD) data which provides job totals associated with both a home Census Block and a work Census Block, Workplace Area Characteristics (WAC) data where jobs are totaled by home Census Block, and Residence Area Characteristics (RAC) where jobs are totaled by work Census Block. We aggregate the commuting flows to Census Tract level.
2. **Training Sample for Share of High-skilled Commuters** — The 2009 National Household Travel Time Survey (NHTS) describes the travel behavior of the American public. We construct a sample of commuting trips during rush hours. We also know the level of education of the travelers. This provides us with a training sample to predict the share of high-skilled workers commuting between a pair of home and work Census Tract based on neighborhood demographic characteristics. We obtain the geocoded version of 2009 NHTS where we know the identities of the home and work Tracts, and merge in neighborhood characteristics from other data sources.
3. **Residential Census Tract Characteristics** — The U.S. Decennial Censuses and 5-year American Community Surveys (ACS) downloaded from The National Historical Geographic Information System (NHGIS) provides the demographic characteristics at the Tract level.
4. **Workplace Census Tract characteristics** — ZIP Code Business Patterns (ZBP) provides annual statistics for businesses with paid employees within the U.S. at the ZIP Code level. We cross-walk ZIP Code level characteristics to the Census Tract level.

## Methodology

Write  $n_{i,j}^s = n_{i,j} \times \mathcal{S}_{i,j}^s$ , where  $n_{i,j}$  is the total number of workers commute between home Tract  $i$  and work Tract  $j$ , and  $\mathcal{S}_{i,j}^s$  is the share of workers who live in  $i$  and work in  $j$ , and belong to skill group  $s$ .  $n_{i,j}$  comes from LODES OD data directly, so it remains to calculate  $\mathcal{S}_{i,j}^s$ . We estimate a predictive model for  $\mathcal{S}_{i,j}^s$  using a sample of nationally representative commuting trips from 2009 NHTS:

$$\mathcal{S}_{i,j}^s = \beta_0 + \mathbf{X}_{i,j}\beta_1 + \varepsilon_{i,j}, \quad (28)$$

where  $\mathbf{X}_{i,j}$  are neighborhood characteristics of the home and work Tracts. The candidate set of  $\mathbf{X}_{i,j}$  includes 25 variables from Census/ACS that describe demographic characteristics of the home Tract, such as log median household income and median home value. We include 7 variables from ZBP that describe workplace characteristics of the work Tract, such as population density and employment density. We include 23 demographic characteristics of the home Tract derived from LODES RAC data such as shares of workers by age, income and industry. We also include

23 workplace characteristics of the work Tract derived from LODES WAC data such as shares of workers by age, income and industry. Last, we include 6 characteristics of the home and work Tract pair from LODES OD data which are shares of workers commuting between the pair of Tracts by age, income and industry sector. We convert all neighborhood characteristics measured in dollar units to be in real 2010 dollars. The full list of candidate explanatory variables are reported in Table A9.

Estimating equation (28) is equivalent to estimating a classification model to predict whether an individual commuter belong to the high-skill group when he commutes between home Tract  $i$  and work Tract  $j$ . We split the NHTS sample into training (80%) and test (20%) samples at the home-work Tract pair level. We train the model using penalized logistic regressions like Lasso, Ridge and Elastic Net, and ensemble methods like Random Forest and XGBoost. For penalized logistic regressions like Lasso, we use 10-fold cross-validation on the test sample to select the optimal penalty parameter. We use the Root Mean Squared Error (RMSE) as the criterion for evaluating predictive accuracy and model selection. The RMSE is defined as

$$\text{RMSE} = \sqrt{\sum_i \sum_j w_{i,j} (\hat{\mathcal{S}}_{i,j}^H - \mathcal{S}_{i,j}^H)^2}$$

where  $\mathcal{S}_{i,j}^H$  is the actual high-skilled ratio of commuters who live in  $i$  and work in  $j$ ,  $\hat{\mathcal{S}}_{i,j}^H$  is the predicted high-skilled ratio from the model, and  $w_{i,j}$  is the weight of the pair of Tracts  $(i, j)$ .<sup>49</sup> We apply personal weights from NHTS in our estimation to ensure the sample of commuters are nationally representative. Table A10 compares the performance of the different models. Our preferred specification is a Lasso Logistic model that achieves the best predictive accuracy in terms of RMSE out-of-sample. The coefficients are reported in A11.

To validate our model, we calculate the high-skilled ratio  $\mathcal{S}_{i,j}^H$  for pairs of Tracts  $(i, j)$  in our training and test samples, and plot them against the model predicted high-skilled ratios  $\hat{\mathcal{S}}_{i,j}^H$  in Figure B31. First, using the full set of predictors, Figures B31a and Figure B31b show the model predicted high-skill ratios agree well with the actual high-skill ratios in both the training and test samples. The slopes of the regression lines are 1.02 and 1.01 with  $R^2$  values of 0.6 and 0.61, respectively. We check the robustness of our model by estimating the model using only variables from LODES so that all information used to predict commute patterns by skill are from the data on commute flows alone. We achieve similar model performance in Figures B31c and Figure B31d, which show the slopes of the regression lines are 1.02 and 1.04 with  $R^2$  values of 0.58 and 0.6, respectively.

We use the estimated model to predict out-of-sample on all pairs of home and work Census Tracts within 50 miles of a high-skilled firm entry site to obtain  $\mathcal{S}_{i,j,t}^s$  for one year prior to the firm entry ( $t = 1$ ), and five years afterwards ( $t = 2$ ), respectively. To do so, we first obtain  $n_{i,j,t}$ , the total number of primary jobs commuting between residential Census Block  $i$  and workplace

---

<sup>49</sup> $w_{i,j}$  is given by the sum of personal weights of commuters who belong to  $(i, j)$ , divided by the sum of personal weights across all commuters.

Census Block  $j$  from LODES OD data for the two periods. We aggregate the commuting flow from Census Block level to Tract level. Next, we prepare the same set of chosen predictors  $\tilde{\mathbf{X}}_{i,j,t}$  from our estimated model from equation (28) for the two periods. Variables measured in dollar units are converted into 2010 real dollars. Finally,  $n_{i,j,t}^s = n_{i,j,t} \times \mathcal{S}_{i,j,t}^s$ .

### C.5.2 Wages by Skill in Workplace Neighborhoods

We calculate the wages by skill group in a workplace Tract  $j$  as a weighted average of the median wage by skill in residential Tracts that have workers commuting to  $j$ , where the weights are given by the number of workers in skill group  $s$  who live in a residential Tract  $i$  and work in  $j$ . Median wage by skill in residential Tract  $i$ ,  $W_{i,t}^s$ , come directly from Census/ACS. Let  $n_{i,j,t}^s$  denote the number of workers of skill group  $s$  who live in residential Tract  $i$  and work in workplace Tract  $j$  in year  $t$ . The average wage by skill in workplace Tract  $j$  is given by:

$$W_{j,t}^s = \sum_i \frac{n_{i,j,t}^s}{\sum_i n_{i,j,t}^s} W_{i,t}^s$$

We crosswalk the wages by skill from Tract to ZIP Code level.

### C.5.3 Rents in Residential Neighborhoods

For each treated residential Tract  $i$  in year  $t$ , we impute a rent index using our house price index constructed using CoreLogic Deeds, together with the house price to rent ratio for its associated entry zone  $k$  calculated from Census/ACS.

Specifically, for each firm entry zone  $k$  in year  $t$ , we calculate an overall house price to rent ratio, using the fractions of renter-occupied housing units in each Census Tract  $i'$  out of all renter-occupied housing units within 30 miles of the entry site as weights. Formally, let  $\text{rhu}_{i',t}$  be the number of renter-occupied housing units in Tract  $i'$  in year  $t$ ; denote the median nominal house price in Census Tract  $i'$  as  $\text{nom\_mhmval}_{i',t}$ , and the median nominal monthly rent in Tract  $i'$  as  $\text{nom\_mrent}_{i',t}$ . The house price to rent ratio in Tract  $i'$  in year  $t$  is equal to  $\frac{\text{nom\_mhmval}_{i',t}}{\text{nom\_mrent}_{i',t} \times 12}$ . Let  $\text{dist}_{i',k}$  be the distance between Tract  $i'$  and the entry site for firm  $k$ . The overall house price to rent ratio for entry zone  $k$  in year  $t$  is thus given by:

$$\text{pr\_ratio}_{k,t} = \sum_{i': \text{dist}_{i',k} \leq 30} \frac{\text{rhu}_{i',t}}{\sum_{i'} \text{rhu}_{i',t}} \times \frac{\text{nom\_mhmval}_{i',t}}{\text{nom\_mrent}_{i',t} \times 12}.$$

Denote the log house price index in the residential neighborhood  $i$  in year  $t$  as  $\text{HPI}_{i,t}$ . The log rent index in Tract  $i$  in entry zone  $k$  in year  $t$  is then given by:

$$r_{i,k,t} = \text{HPI}_{i,t} - \log(\text{pr\_ratio}_{k,t}) - \log(12).$$

We crosswalk the rent index from Tract to ZIP Code level.

#### C.5.4 Initial Distribution of Birthplace by Skill and Ownership

We obtain the distribution of workers by skill and homeownership status within each city of firm entry from 2000 Census. This distribution is used to generate model predicted choice probabilities in order to match observed choice probabilities in our estimation procedure.

#### C.5.5 Crosswalk of Neighborhood Characteristics from Tract to ZCTA level

We crosswalk a couple neighborhood characteristics constructed at the Census Tract level to ZCTA level to be used for estimating our model. We use 2010 ZCTA boundaries instead of USPS ZIP Codes because the former are standardized by the Census Bureau and their boundaries are fixed over time. Details are provided below.

##### Crosswalk of Rents and Wages by Skill

Suppose ZCTA 1 is fully covered by  $m$  Tracts indexed by  $t_{1,1}, t_{1,2}, \dots, t_{1,m}$ , and let the share of ZCTA 1's population living in the  $m$  Tracts indexed by  $p_{1,1}, p_{1,2}, \dots, p_{1,m}$ , which sum up to 1. We have wages and rents measured at the Tract level. Therefore, the population shares can be treated as weights to calculate wages and rents at the ZCTA level. For example, the wage by skill  $W_1^s$  in ZCTA 1 is given by:

$$W_1^s = \sum_i W_{1,i}^s \times p_{1i}.$$

The crosswalk of rents follows the same procedure.

##### Crosswalk of Driving Time

The goal is to calculate the driving time between ZCTA 1 and 2,  $d_{1,2}$ . Suppose ZCTA 1 is fully covered by  $m$  Tracts indexed by  $t_{1,1}, t_{1,2}, \dots, t_{1,m}$ , and let the share of ZCTA 1's population living in the  $m$  Tracts indexed by  $p_{1,1}, p_{1,2}, \dots, p_{1,m}$ , which sum up to 1. Suppose ZCTA 2 is fully covered by  $n$  Tracts indexed by  $t_{2,1}, t_{2,2}, \dots, t_{2,n}$ , and let the share of ZCTA 2's population living in the  $n$  Tracts indexed by  $p_{2,1}, p_{2,2}, \dots, p_{2,n}$ , which sum up to 1. We have the driving time between Tract  $i$  associated with ZCTA 1 and Tract  $j$  associated with ZCTA 2. Denoted the pair of Tracts as  $(t_{1,i}, t_{2,j})$ . Denote the driving time between  $t_{1,i}$  and  $t_{2,j}$  as  $d_{1i,2j}$ .

We calculate the driving time between ZCTA 1 and 2 as a weighted average of the driving time between pairs of Tracts associated with the two ZCTAs. We construct the weight for each pair of tracts  $(t_{1,i}, t_{2,j})$  as  $p_{1,i} \times p_{2,j}$ , that is the product of the share of ZCTA 1's population living in Tract  $i$  and share of ZCTA 2's population living in Tract  $j$ . The matrix of weights can be written as:

$$\begin{pmatrix} p_{1,1}p_{2,1} & \dots & p_{1,1}p_{2,n} \\ \dots & p_{1,i}p_{2,j} & \dots \\ p_{1,m}p_{2,1} & \dots & p_{1,m}p_{2,n} \end{pmatrix},$$

where row  $i$  represents Tract  $t_{1,i}$  associated with ZCTA 1, and column  $j$  represents Tract  $t_{2,j}$  associated with ZCTA 2. Note the weights in the matrix sum up to 1. The driving time  $d_{1,2}$  is

given by the weighted average:

$$d_{1,2} = \sum_i^m \sum_j^n d_{1i,2j} \times p_{1,i} \times p_{2,j}.$$

### Crosswalk of Choice Probabilities

The goal is to calculate the share of population in skill group  $s$  who choose residential ZCTA 1 and workplace ZCTA 2,  $\pi_{1,2}^s$ . Suppose ZCTA 1 is fully covered by  $m$  Tracts indexed by  $t_{1,1}, t_{1,2}, \dots, t_{1,m}$ , and let the share of each Tract's population living in ZCTA 1 indexed by  $tr_{1,1}, tr_{1,2}, \dots, tr_{1,m}$ . Suppose ZCTA 2 is fully covered by  $n$  Tracts indexed by  $t_{2,1}, t_{2,2}, \dots, t_{2,n}$ , and let the share of each Tract's population living in ZCTA 2 indexed by  $tr_{2,1}, tr_{2,2}, \dots, tr_{2,n}$ . We have the share of population of skill group  $s$  who choose residential Tract  $i$  associated with ZCTA 1 and workplace Tract  $j$  associated with ZCTA 2,  $\pi_{1i,2j}^s$ .

We can derive that for the pair of Tracts,  $(t_{1i}, t_{2,j})$ , its area inside the pairs of residential ZCTA 1 and workplace ZCTA 2 contains a population of  $N_k^s \times \pi_{1i,2j}^s \times tr_{1,i} \times tr_{2,j}$ , where  $N_k^s$  is the population of skill group  $s$  living within 50 miles from entry site of firm  $k$ . Hence,  $\pi_{1,2}^s$  is given by:

$$\pi_{1,2}^s = \sum_i^m \sum_j^n \pi_{1i,2j}^s \times tr_{1,i} \times tr_{2,j}.$$

## C.6 10 Cases of Firm Entry

Due to concerns that other large firms are opening right around the time the firms in our sample are opening, we randomly select 10 cases from the firm entry sample to understand them more in detail. We find that for only 2 of the 10 cases, it looks like another large firm entered just before, or there was generally a lot of economic development existed prior to firm entry. For 2 to 3 of them, more research is needed. For 5 to 6 of them, we did not find any evidence of other large firm entries around the same time.

### 1. Micron in Lehi, Utah in 1995

- Micron, a microchip plant, opened in Lehi, Utah in 1995. The area is now nicknamed "Silicon Slopes" due to the many tech companies locating in that area. However, at the time that Micron opened in 1995, there appears to be no other large high-skilled firms in the vicinity. Adobe opened a large office about 10 minutes away in 2012, which is well after Micron's entry. Other companies including a hospital opened in June 2015, and Ancestry.com moved there in 2016.
- Takeaway: there's obviously a lot of economic development happening in the area, but most of it seems well after 1995

### 2. Specialized Devices in Moorpark, CA in 1998

- This company became big, like 700 people (<https://www.latimes.com/archives/la-xpm-2000-sep-02-me-14401-story.html>), and then small again, like 200, but it's large compared with the local population ([https://en.wikipedia.org/wiki/Moorpark,\\_California#Economy](https://en.wikipedia.org/wiki/Moorpark,_California#Economy)). PennyMac Loan Services is large and close by, but it was founded in 2008. Generally, it's hard to find anything online about this company or the city and its economy in 1998.
  - Takeaway: hard to tell due to lack of information, but probably fine because it's large relative to the surroundings
3. Caris Diagnostics in Irving, TX in 2008
- Caris is in the Las Colinas neighborhood of Irving, TX. According to Wikipedia, "During the 1980s building boom, Las Colinas became a popular location for relocating companies and office developers, attracting many corporations, including the global headquarters of four Fortune 500 companies and offices of more than 30 others, such as ExxonMobil, GTE Telephone (now Verizon), Kimberly-Clark and Associates Corp." ([https://en.wikipedia.org/wiki/Las\\_Colinas](https://en.wikipedia.org/wiki/Las_Colinas))
  - Takeaway: economic development probably happened first, then Caris moved in
4. LegalZoom in Austin, TX in 2010
- Seems like some of the jobs are not high tech: "Austin office will include sales positions, order fulfillment, customer service and technical support representatives."
  - Takeaway: likely not much going on prior to firm entry, need to do more research to confirm
5. Northrop Corp in Palmdale, CA in 1992
- Takeaways: it is in the aerospace industry so it and several other companies are very close to the airport. Seems like Lockheed entered in 1989, and Northrop entered in 1992.
6. Edison Chouest in Houma, LA in 2008
- Takeaways: It is in the shipping industry so it has to be by the water. Otherwise difficult to pinpoint the precise location as it has since moved. Need to do more research.
7. BMW in Greer, SC in 1992
- Takeaways: huge employer that entered early so unlikely that there were other firms existing at the time.
8. River Hill Coal in Karthaus, PA in 2003
- Takeaways: Nothing else in the vicinity. It is a coal mining facility.
9. General Motors in Wentzville, MO in 1996



- Takeaways: GM is biggest employer by far. Factory actually opened in 1983, and there was an expansion in 1996. This site is a car manufacturing facility.

10. Comcast Cable in Newark, DE in 2007 or 2010

- Takeaways: Seems like a call center so might have some jobs that are not high-skilled. Not completely sure that we have identified the correct entry.

## D Propensity Score Model

Using our sample of 410 firm entry announcements<sup>50</sup>, we build a propensity score model that captures the likelihood that a particular ZIP Code is suitable for the entry of a large high-skilled firm. The full sample consists of all ZIP Codes within 50 miles of the entry site, and a ZIP Code is assigned an indicator equal to 1 if it was ever home to one of the 410 entries and 0 otherwise.<sup>51</sup> Using the ex ante 1990 characteristics of each ZIP Code and its surrounding ZIP Codes prior to all firm entries in our sample, we estimate a regression of the conditional probability that a ZIP Code receives a high-skilled firm entry:

$$\mathbb{E}[Y_z|N_{r,z}, W_{r,z}, D_z] = \text{logistic} \left( \beta_0 + \sum_r \beta_{1r} N_{r,z} + \sum_r \beta_{2r} W_{r,z} + \beta_3 D_z + \sum_r \beta_{4r} N_{r,z} \times D_z \right) \quad (29)$$

where  $Y_z$  is an indicator that equals 1 if the ZIP Code ever has one of the 410 entries;  $N_{r,z}$  are neighborhood demographic characteristics in successive rings  $r$  around ZIP Code  $z$ ,  $W_{r,z}$  are workplace characteristics in successive rings  $r$  around ZIP Code  $z$ ,  $D_z$  represents two variables: (1) the distance of ZIP Code  $z$  to the closest Central Business District (CBD) which measures a neighborhood’s urban vs. rural status (2) The employment-to-population ratio in ZIP Code  $z$  which measures how commercial vs. residential a neighborhood is, and  $N_{r,z} \times D_z$  are the interaction effects between demographic characteristics in rings  $r$  around  $z$  with the distance of ZIP Code  $z$  to the closest CBD and the employment-to-population ratio in ZIP Code  $z$ , respectively. The rings  $r$  are defined as being within 0–5 miles, 5–10 miles, 10–20 miles, and 20–30 miles of ZIP Code  $z$ .  $W_{r,z}$  and  $D_z$  proxy for access to jobs in nearby workplaces for workers living in ZIP Code  $z$ , and  $N_{r,z}$  reflect the demographic composition of residents in ZIP Code  $z$  and its neighboring areas. Table A16 and Table A17 contain the complete list of demographic and workplace variables in the model.

We estimate equation 29 using a LASSO-Logistic regression. ZIP Codes associated with 90% of the 410 firms comprise the training data, and ZIP Codes associated with 10% of the 410 firms are held out as the testing data. LASSO regularization avoids overfit by penalizing the number of variables in the model by adding a term to the cost function of a conventional logistic regressions, which represents the sum of the magnitudes of all the coefficients in the model. Table A5 shows the estimated parameters of the propensity score model. Noticeably, the probability of receiving a

<sup>50</sup>The sample consists of 391 actual entries and 19 announced entries that never occurred ex post. The main analysis uses only the 391 entries that did occur.

<sup>51</sup>ZIP Codes are defined by 2010 ZIP Code Tabulation Areas (ZCTAs), and are time-invariant in their boundaries.

high-skilled firm entry in a ZIP Code is positively correlated with nearby job densities in high-tech industries, employment to population ratio which measures whether an area is more commercial vs. residential, and variables that indicate the local workforce is young.

We validate of the propensity score model by comparing the job growth in high-tech industries in the treatment group of ZIP Codes that actually received any of the firm entries from our sample of 391 firms with job growth in the control group of ZIP Codes that did not receive an entry from our sample, but are predicted to have a high probability of receiving a high-skilled firm entry by our model. Our sample of firm entries does not capture of universe of large high-skilled firm entries. The goal here is to examine whether the treated ZIP Codes indeed experienced significantly higher high-skilled job growth relative to the control ZIP Codes, which are the potential sites for entry. If the control group received a comparable amount of high-skilled firm entries, then the treatment effect estimated from equation (1) is likely to be biased downward. We estimate the following regression which is similar to (1) for neighborhood level outcomes, except the treatment and control groups are already similar in terms of observables by design.<sup>52</sup>

$$Y_{i,k,t} = \alpha_0 + \sum_{\tau} \delta_{\tau} D_{\tau,k} \times g_{i,k} + \eta_{i,k} + \zeta_{k,t} + \varepsilon_{i,k,t} \quad (30)$$

where  $Y_{i,k,t}$  is the number of establishment of a certain size in ZIP Code  $i$  in the metropolitan area defined as a 50-mile ring around the entry site  $k$  in year  $t$ , measured from ZBP.  $D_{\tau,k}$  is an indicator for  $\tau$  years relative to the announcement year, and  $g_{i,k}$  is an indicator for whether ZIP Code  $i$  in the metropolitan area  $k$  is in the treatment group of ZIP Codes that actually received an entry in our sample of 391 firm entries. In addition,  $\eta_{i,k}$  are a set of metropolitan area by ZIP fixed effects, and  $\zeta_{k,t}$  are a set of metropolitan area by year fixed effects. Our coefficient of interest, quantifying the relative difference in high-tech job growths between the treatment and control groups, is denoted by  $\delta_{\tau}$ .

Figure 3 plots the event study coefficients. It shows there are parallel trends between our treatment and control groups in the number of high-tech establishments of size 500+ in panel (a) and of size 100–499 in panel (b) before the firm entry announcement. After the entry, we observe a significantly higher growth of high-tech jobs at both large and smaller establishments. In Table A6, Panel A, we summarize the treatment effects 6 to 10 years after the firm entry announcement. Column (2) of Table A6 shows that the number of high-tech establishments of size 500+ measured from the ZBP in the control ZIP Codes on average increased by 0.11 after the entry announcement. The treated ZIP Codes on average experience an additional increase of 0.75 high-tech establishments of size 500+ 6 to 10 years after announcement.<sup>53</sup> Column (1) of Table A6 shows that relative to

<sup>52</sup>Using Figure 2 as an illustration, for the case of Pfizer’s entry, we are basically comparing the high-skilled job growth received by the ZIP Code marked by the blue dot in Ann Arbor with other red ZIP Codes on the map within the same metropolitan area, after the entry announcement. We do this comparison by pooling all the entries in our sample.

<sup>53</sup>This implies the treated ZIP Codes on average experience an increase of 0.86 high-tech establishments of size 500+ 6–10 years after announcement. If all 391 firm entries in our sample actually happened and reached their target capacity of 500+, we should expect an increase of exactly 1. In reality, all the firm entries in our sample did happen but some of them never reached their target capacity. There were also firm exits. Note we only observe the total

the control ZIP Codes, the treated ZIP Codes on average experience an additional increase of 2.48 high-tech establishments of size between 100 and 499. This suggests the large firm entries in our sample also brought in significantly more high-tech jobs from smaller and complementary firms following our large firm entries as well, relative to the control group which are other desirable places for high-skilled firm entry in the same metropolitan area. In columns (1) and (2), the control ZIP Codes are those that did not received an entry in our sample, but are from the top quintile of ZIP Codes within 50 miles of the firm entries in our sample, that have the highest probability of receiving a high-skilled firm entry predicted by our model. The results are robust to using the top decile of ZIP Codes in columns (3) and (4).

## E Model

### E.1 Worker's Choice for Residential and Workplace Neighborhoods

A worker  $\omega$ , who belongs to skill group  $s \in \{H, L\}$  and homeownership group  $o \in \{O, R\}$  and has chosen to live in neighborhood  $i \in \mathcal{N}$  and work in neighborhood  $j \in \mathcal{N}$  in period  $t$  has the indirect utility over a pair of locations  $(i, j)$  given by:

$$V_{i,j,\omega,t}^{s,o} = w_{j,t}^s - \theta^o r_{i,t} - \lambda^s \tau_{i,j} + a_{i,j,t}^s - b_{\omega}^s \mathbf{1}_{\omega,t}(i \neq i_0) + \sigma^s \varepsilon_{i,j,\omega,t} \quad (31)$$

- $w_{j,t}^s$ : log wage in workplace neighborhood  $j$
- $\theta^o$ : share of income spent on the local good. In the full model, we allow it to be potentially different for renters and owners to account for their differences in preferences.
- $r_{i,t}$ : log rent in residential neighborhood  $i$
- $\tau_{i,j}$ : commuting time from  $i$  to  $j$
- $a_{i,j,t}^s$ : amenity for  $(i, j)$  which is skill specific
- $b_{\omega}^s \mathbf{1}_{\omega,t}(i \neq i_0)$ : birthplace dis-amenity
- $\sigma^s \varepsilon_{i,j,\omega,t}$ : idiosyncratic preference distributed T1EV with variance  $\sigma^s$ .

Here we define  $v_{i,j,\omega,t}^{s,o} \equiv w_{j,t}^s - \theta^o r_{i,t} - \lambda^s \tau_{i,j} + a_{i,j,t}^s - b_{\omega}^s \mathbf{1}_{\omega,t}(i \neq i_0)$ . The setup is the conditional logit model first formulated in the utility maximization context by [McFadden \(1973\)](#). The total expected population of workers choosing a pair  $(i, j)$  is simply the probability of each worker choosing  $(i, j)$ , summed over all workers.<sup>54</sup> Thus, the total populations of the four demographic

---

number of establishments by size in ZBP and not the flow, so there is no way to measure new firm entries only.

<sup>54</sup>By the T1EV assumption of the idiosyncratic preference, the probability worker  $\omega$  of skill group  $s$  chooses to live

groups living in  $i$  and working in  $j$  are:

$$HO_{i,j,t} = \sum_{\omega \in \mathcal{HO}_t} \frac{\exp\left(v_{i,j,\omega,t}^{H,O}/\sigma^H\right)}{1 + \sum_{i'} \sum_{j'} \exp\left(v_{i',j',\omega,t}^{H,O}/\sigma^H\right)}, \quad (33)$$

$$HR_{i,j,t} = \sum_{\omega \in \mathcal{HR}_t} \frac{\exp\left(v_{i,j,\omega,t}^{H,R}/\sigma^H\right)}{1 + \sum_{i'} \sum_{j'} \exp\left(v_{i',j',\omega,t}^{H,R}/\sigma^H\right)}, \quad (34)$$

$$LO_{i,j,t} = \sum_{\omega \in \mathcal{LO}_t} \frac{\exp\left(v_{i,j,\omega,t}^{L,O}/\sigma^L\right)}{1 + \sum_{i'} \sum_{j'} \exp\left(v_{i',j',\omega,t}^{L,O}/\sigma^L\right)}, \quad (35)$$

$$LR_{i,j,t} = \sum_{\omega \in \mathcal{LR}_t} \frac{\exp\left(v_{i,j,\omega,t}^{L,R}/\sigma^L\right)}{1 + \sum_{i'} \sum_{j'} \exp\left(v_{i',j',\omega,t}^{L,R}/\sigma^L\right)}. \quad (36)$$

where  $\mathcal{HO}_t$ ,  $\mathcal{HR}_t$ ,  $\mathcal{LO}_t$  and  $\mathcal{LR}_t$  denote the set of workers of each demographic group in period  $t$ , respectively.

## E.2 Labor Supply

Labor supply to each neighborhood follows from the model of individual residential and workplace neighborhood choice. For each skill group  $s \in \{H, L\}$  and workplace neighborhood  $j$ , labor supply is the sum of all workers who choose to work there, regardless of the residential neighborhood  $i$  they choose to live in:

$$H_{j,t} = \sum_i (HO_{i,j,t} + HR_{i,j,t}), \quad (37)$$

$$L_{j,t} = \sum_i (LO_{i,j,t} + LR_{i,j,t}). \quad (38)$$

## E.3 Labor Demand

Each workplace neighborhood  $j$  has homogeneous firms, indexed by  $d$ . We assume high and low-skilled workers in the same neighborhood work in the same set of firms. We assume the labor market is perfectly competitive. The firm's production function is Cobb-Douglas with constant returns to scale in aggregate labor input  $N_{d,j,t}$  and capital  $K_{d,j,t}$ . The aggregate labor hired by each firm combines high-skilled and low-skilled labor as imperfect substitutes into the production

in neighborhood  $i$  and work in neighborhood  $j$  in period  $t$  is:

$$\pi_{i,j,\omega,t}^{s,o} = \Pr\left(V_{i,j,\omega,t}^{s,o} = \max_{i',j'} V_{i',j',\omega,t}^{s,o}\right) = \frac{\exp\left(\frac{v_{i,j,\omega,t}^{s,o}}{\sigma_s}\right)}{1 + \sum_{i'} \sum_{j'} \exp\left(\frac{v_{i',j',\omega,t}^{s,o}}{\sigma_s}\right)}. \quad (32)$$

of a homogeneous tradeable good. Firms' production function is:

$$Y_{d,j,t} = N_{d,j,t}^\alpha K_{d,j,t}^{1-\alpha}, \quad (39)$$

$$N_{d,j,t} = \left( X_{j,t}^L (L_{d,j,t})^\rho + X_{j,t}^H (H_{d,j,t})^\rho \right)^{\frac{1}{\rho}}. \quad (40)$$

where  $\alpha$  is the share of labor,  $X_{j,t}^s$  is a skill and neighborhood specific productivity shifter, and  $\frac{1}{1-\rho}$  is the elasticity of labor substitution. Since the labor market is perfectly competitive, firms set wages equal to the marginal product of labor. A frictionless capital market supplies capital perfectly elastically at price  $\kappa_t$ , which is constant across the nation<sup>55</sup>. Each firm's (inverse) demand for labor is:

$$\begin{aligned} W_{d,j,t}^H &= \alpha N_{d,j,t}^{\alpha-\rho} K_{d,j,t}^{1-\alpha} H_{d,j,t}^{\rho-1} X_{j,t}^H, \\ W_{d,j,t}^L &= \alpha N_{d,j,t}^{\alpha-\rho} K_{d,j,t}^{1-\alpha} L_{d,j,t}^{\rho-1} X_{j,t}^H. \end{aligned}$$

Firm-level labor demand translates directly into neighborhood-level aggregate demand since firms face constant return to scale production function and identical production technology:

$$\begin{aligned} W_{j,t}^H &= \alpha N_{j,t}^{\alpha-\rho} K_{j,t}^{1-\alpha} H_{j,t}^{\rho-1} X_{j,t}^H, \\ W_{j,t}^L &= \alpha N_{j,t}^{\alpha-\rho} K_{j,t}^{1-\alpha} L_{j,t}^{\rho-1} X_{j,t}^H. \end{aligned}$$

Plugging in the price of capital into the (inverse) labor demand function and rearranging yields:

$$w_{j,t}^H = \ln W_{j,t}^H = c_t + (1-\rho) \ln N_{j,t} + (\rho-1) \ln H_{j,t} + \ln X_{j,t}^H, \quad (41)$$

$$w_{j,t}^L = \ln W_{j,t}^L = c_t + (1-\rho) \ln N_{j,t} + (\rho-1) \ln L_{j,t} + \ln X_{j,t}^L. \quad (42)$$

where  $c_t = \ln \alpha + \frac{1-\alpha}{\alpha} \ln(\frac{1-\alpha}{\kappa_t})$  is a constant.

## E.4 Housing Supply

Housing supply is upward sloping in each neighborhood to capture features of city neighborhoods such as zoning restrictions and congestion.

$$r_{i,t} = \iota_t + k_i \ln(HD_{i,t}), \quad (43)$$

where  $k_i$  characterizes the elasticity of the supply of housing in neighborhood  $i$ , and  $HD_{i,t}$  is aggregate housing demand in neighborhood  $i$ . We assume that this parameter is exogenously determined by geography and local land use regulations.

<sup>55</sup>In equilibrium, the marginal product of capital is equal to the price of capital:

$$\kappa_t = (1-\alpha) N_{j,t}^\alpha K_{j,t}^{-\alpha}.$$

## E.5 Housing Demand

Housing demand in each residential neighborhood  $i$  is the aggregate housing demand by either skill and homeownership group choosing to reside there, regardless of where they work:

$$HD_{i,t} = \theta^O \sum_j \left( HO_{i,j,t} \frac{W_{j,t}^H}{R_{i,t}} + LO_{i,j,t} \frac{W_{j,t}^L}{R_{i,t}} \right) + \theta^R \sum_j \left( HR_{i,j,t} \frac{W_{j,t}^H}{R_{i,t}} + LR_{i,j,t} \frac{W_{j,t}^L}{R_{i,t}} \right). \quad (44)$$

## E.6 Equilibrium

Equilibrium in this model is defined by a menu of wages and rents, Equilibrium in this model is defined by a menu of wages and rents,  $(w_{j,t}^{H*}, w_{j,t}^{L*}, r_{i,t}^*)$ , with populations  $(HO_{i,j,t}^*, HR_{i,j,t}^*, LO_{i,j,t}^*, LR_{i,j,t}^*)$  such that:

The high-skilled labor demand equals high-skilled labor supply:

$$H_{j,t}^* = \sum_i (HO_{i,j,t}^* + HR_{i,j,t}^*), \quad (45)$$

$$= \sum_i \left( \sum_{\omega \in \mathcal{H}\mathcal{O}_t} \frac{\exp(v_{i,j,\omega,t}^{H,O}/\sigma^H)}{1 + \sum_{i'} \sum_{j'} \exp(v_{i',j',\omega,t}^{H,O}/\sigma^H)} + \sum_{\omega \in \mathcal{H}\mathcal{R}_t} \frac{\exp(v_{i,j,\omega,t}^{H,R}/\sigma^H)}{1 + \sum_{i'} \sum_{j'} \exp(v_{i',j',\omega,t}^{H,R}/\sigma^H)} \right). \quad (46)$$

$$w_{j,t}^{H*} = c_t + (1 - \rho) \ln N_{j,t}^* + (\rho - 1) \ln H_{j,t}^* + \ln X^H. \quad (47)$$

The low-skilled labor demand equals low-skilled labor supply:

$$L_{j,t}^* = \sum_i (LO_{i,j,t}^* + LR_{i,j,t}^*), \quad (48)$$

$$= \sum_i \left( \sum_{\omega \in \mathcal{L}\mathcal{O}_t} \frac{\exp(v_{i,j,\omega,t}^{L,O}/\sigma^L)}{1 + \sum_{i'} \sum_{j'} \exp(v_{i',j',\omega,t}^{L,O}/\sigma^L)} + \sum_{\omega \in \mathcal{L}\mathcal{R}_t} \frac{\exp(v_{i,j,\omega,t}^{L,R}/\sigma^L)}{1 + \sum_{i'} \sum_{j'} \exp(v_{i',j',\omega,t}^{L,R}/\sigma^L)} \right). \quad (49)$$

$$w_{j,t}^{L*} = c_t + (1 - \rho) \ln N_{j,t}^* + (\rho - 1) \ln L_{j,t}^* + \ln X^L. \quad (50)$$

Housing demand equals housing supply:

$$r_{i,t}^* = \iota_t + k_i \ln (HD_{i,t}^*) \quad (51)$$

$$HD_{i,t}^* = \theta^O \sum_j \left( HO_{i,j,t}^* \frac{\exp(w_{j,t}^{H*})}{\exp(r_{i,t}^*)} + LO_{i,j,t}^* \frac{\exp(w_{j,t}^{L*})}{\exp(r_{i,t}^*)} \right) + \theta^R \sum_j \left( HR_{i,j,t}^* \frac{\exp(w_{j,t}^{H*})}{\exp(r_{i,t}^*)} + LR_{i,j,t}^* \frac{\exp(w_{j,t}^{L*})}{\exp(r_{i,t}^*)} \right) \quad (52)$$

## E.7 Welfare Incidence of a High-Skilled Firm Entry

We can write the expected utility of workers from demographic group  $\{s, o\}$  in period  $t$  as:

$$\begin{aligned} \mathcal{V}_t^{s,o} &= \mathbb{E}_\varepsilon \left[ \max_{i,j} \{v_{i,j,\omega,t}^{s,o} + \sigma^s \varepsilon_{i,j,\omega,t}\} \right] \\ &= \sigma_s \sum_{i_0 \in \mathcal{N}} \pi_{i_0}^{s,o} \ln \left( 1 + \sum_i \sum_j \exp \left( \frac{v_{i,j,\omega,t}^{s,o,i_0}}{\sigma_s} \right) \right), \end{aligned}$$

where  $\pi_{i_0}^{s,o}$  denotes the share of workers from group  $\{s, o\}$  born in neighborhood  $i_0$ . The expectation operator  $\mathbb{E}_\varepsilon$  is defined over the idiosyncratic preference term  $\varepsilon_{i,j,\omega,t}$ . The last line follows from the assumption that preference heterogeneity is drawn from a T1EV distribution and we can integrate out over the error distribution.  $v_{i,j,\omega,t}^{s,o,i_0}$  denotes the mean utility of the option  $(i, j)$  in period  $t$  for workers of type  $\{s, o, i_0\}$ .

To approximate the scenario of a high-skilled firm entry into a neighborhood, consider the case where the productivity of high-skilled workers increases relative to the productivity of low-skilled workers in the area where the firm enters. To derive the first-order welfare incidence of a high-skilled firm entry, we take the derivative of a worker's expected welfare with respect to the productivity shock associated with a high-skilled firm entry:<sup>56</sup>

$$\frac{d\mathcal{V}_t^{s,o}}{dX} = \sum_{i_0 \in \mathcal{N}} \pi_{i_0}^{s,o} \sum_i \sum_j \frac{\partial \mathcal{V}_t^{s,o}}{\partial v_{i,j,\omega,t}^{s,o,i_0}} \frac{dv_{i,j,\omega,t}^{s,o,i_0}}{dX} \quad (53)$$

$$= \sum_{i_0 \in \mathcal{N}} \pi_{i_0}^{s,o} \sum_i \sum_j \pi_{i,j,t}^{s,o,i_0} \left( \frac{dw_{j,t}^s}{dX} - \theta^o \frac{dr_{i,t}}{dX} - \underbrace{\lambda^s \frac{d\tau_{i,j}}{dX} + \frac{da_{i,j,t}^s}{dX} - \frac{db_\omega \mathbf{1}_{\omega,t}(i \neq i_0)}{dX}}_{=0} \right) \quad (54)$$

where  $\pi_{i,j,t}^{s,o,i_0}$  is the share of workers of type  $\{s, o, i_0\}$  who choose the option  $(i, j)$  in period  $t$  prior to the shock.  $dX$  represents a positive, localized, skill-biased productivity shock. The incidence of the high-skilled firm entry may be approximated by the change in prices (wages and rents) experienced by the workers weighted by the share of them born in  $i_0$ , and their choice probabilities of choosing each option  $(i, j)$  before the shock. Of course, after the shock, some workers who were initially “marginal” workers between choosing  $(i, j)$  and  $(i', j')$  will re-optimize and move. However, the Envelope Theorem suggests agents will only have small gains in utility by switching choices in response to marginal changes in prices. Hence, these behavioral effects due to price changes could only have secondary effects on worker welfare. Otherwise, agents would not be optimizing. Of course, a high-skilled firm entry might not be a marginal change.

## E.8 Simulation for a High-Skilled Firm Entry

We perform a simulation of the effect of introducing a skill-biased productivity shock in one neighborhood on equilibrium wages, rents, residential and workplace neighborhood choices, and worker's welfare. The spacial distribution of neighborhoods are similar to those in the city of Seattle, WA. A high-skilled productivity shock happens in a census tract in downtown. We assume the strength of the productivity shock dissipates across space following a standard normal density function. A low-skilled productivity shock of half the strength also happens downtown due to spillover from the high-skilled sector to low-skilled sector of industries. The low-skilled productivity shock dissipates

---

<sup>56</sup>The relationship  $\frac{\partial \mathcal{V}_t^{s,o}}{\partial v_{i,j,\omega,t}^{s,o,i_0}} = \pi_{i,j,t}^{s,o,i_0}$  follows directly from assuming that preference heterogeneity is drawn from a T1EV distribution (Train (2009)). However, this relationship also holds independent of the distribution of the taste heterogeneity. See Busso et al. (2013).

across space as well. In period 0, 2500 high-skilled and 2500 low-skilled workers are assigned to each neighborhood. The population born outside is set to be 10 times of the initial population inside the city. We make the parameter choices for the model's calibration in Table A8.

## E.9 Estimation Procedure

Workers choose the pair of home-work locations  $(i, j)$  that maximizes utility. In this model, the share of incumbent residents in skill group  $s$  that choose  $(i, j)$  is given by:

$$\pi_{i,j,t}^s = \int \pi_{i,j,\omega,t}^{s,i_0} dF^s(\mathbf{i}_0), \quad (55)$$

$$\int \frac{\exp\left(\delta_{i,j,t}^s - b_\omega \mathbf{1}_{\omega,t}(i \neq i_0)\right)}{1 + \sum_{i'} \sum_{j'} \exp\left(\delta_{i',j',t}^s - b_\omega \mathbf{1}_{\omega,t}(i' \neq i_0)\right)} dF^s(\mathbf{i}_0), \quad (56)$$

where  $F^s(\mathbf{i}_0)$  is the incumbents' distribution of "birthplace", i.e., initial locations in period 0, for skill group  $s$ .

To estimate the model, we follow a strategy similar to Berry (1994) and Berry et al. (1995), complemented with micro-moments from Infutor migration data, similar to Petrin (2002). Here we summarize key details of the procedure:

- From the initial spatial distribution of where high- and low-skilled incumbent residents are living at in period 0 (from 2000 Census), we assign each incumbent resident to his "birthplace" neighborhood. Let  $N_{i_0} \equiv N_{i_0}^H + N_{i_0}^L$  be the incumbent residents born in residential neighborhood  $i_0 \in \mathcal{N}$ , where  $N_{i_0}^H$  and  $N_{i_0}^L$  are the number of high- and low-skilled incumbents born in neighborhood  $i_0$ , respectively.  $N^s = \sum_{i_0} N_{i_0}^s$  is the total number of incumbents in skill group  $s$ .
- Given the initial distribution of incumbents and a vector of  $\delta_{i,j,t}^s$ , we can calculate the model predicted share of incumbents choosing each pair of home-work locations  $(i, j)$  by adding up the simulation-specific logit choice probabilities, i.e.,

$$\hat{\pi}_{i,j,t}^s(\delta(\theta), \theta) = \frac{1}{N^s} \sum_{i_0 \in \mathcal{N}} \sum_{\omega}^{N_{i_0}^s} \hat{\pi}_{i,j,\omega,t}^{s,i_0}(\delta(\theta), \theta), \quad (57)$$

$$= \frac{1}{N^s} \sum_{i_0 \in \mathcal{N}} \sum_{\omega}^{N_{i_0}^s} \frac{\exp\left(\delta_{i,j,t}^s(\theta) - b_\omega \mathbf{1}_{\omega,t}(i \neq i_0)\right)}{1 + \sum_{i'} \sum_{j'} \exp\left(\delta_{i',j',t}^s(\theta) - b_\omega \mathbf{1}_{\omega,t}(i' \neq i_0)\right)}, \quad (58)$$

where  $\theta$  is the vector of model parameters and  $\delta^s(\theta)$  is the vector of mean utilities of choices  $(i, j)$  for skill group  $s$ .

- Iterate the equation  $\delta^{h+1} = \delta^h + (\ln \pi - \ln \hat{\pi}(\delta(\theta), \theta))$ ,  $h = 0, \dots, H$  until it converges, where  $\pi$  are the observed vector of shares of choices and  $\hat{\pi}(\delta(\theta), \theta)$  are the vector of model predicted



shares of choices,  $H$  is the smallest integer such that  $\|\delta^H - \delta^{H-1}\|$  is smaller than some tolerance level, and  $\delta^H$  is the approximation to  $\delta$ . It is guaranteed to converge because it is a contraction mapping proven by [Berry et al. \(1995\)](#). We start with some initial values for  $\delta$  such as those given by a standard logit model,  $\delta_{i,j,t}^s = \ln \pi_{i,j,t}^s - \ln \pi_{0,t}^s$ , where  $\pi_{0,t}^s$  is the share of incumbents of skill group  $s$  who choose to live or work outside the entry zone.

- Given a guess of  $b_\omega$ , we can solve for the  $\delta_{i,j,t}^s$  that most closely match the observed choice probabilities.
- We then decompose the mean utility estimates into how incumbent residents value wage, rent and amenity, and take the first difference of the mean utility estimates across period 1 and 2:

$$\Delta \delta_{i,j}^s = \underbrace{\beta^s (\Delta w_j^s - \theta^R \Delta r_i)}_{\text{Changes in real wages: } \equiv \Delta \tilde{w}_{i,j}^s} + \phi_{h^1, h^2, j}^s + \Delta \xi_{i,j}^s. \quad (59)$$

Note the commuting cost term drops out. we add a set of fixed effects to the first-difference equation for each skill group  $s$ , denoted by  $\phi_{h^1, h^2, j}^s$ . These are distance bins of residential neighborhood  $i$  to nearby potential sites for high-skilled firm entry, that we used in our reduced-form strategy in [3.2](#), interacted with the entry zone  $k$  and the workplace neighborhood  $j$ .<sup>57</sup> As econometricians, we observe changes in wages in workplace neighborhoods,  $\Delta w_j^s$ , and changes of rents in residential neighborhoods,  $\Delta r_i$ . However, we do not observe  $\Delta \xi_{i,j}^s \equiv \beta^s \Delta a_{i,j}^s$ , which is the change in exogenous amenities across the two periods for the pair  $(i, j)$  for workers of skill group  $s$ .

- The error term in equation (59) can be written as:

$$\Delta \xi_{i,j}^s = \Delta \delta_{i,j}^s(\pi, b) - \phi_{h^1, h^2, j}^s - \beta^s \Delta \tilde{w}_{i,j}^s \quad (60)$$

Note it is the observed market shares,  $\pi$ , that enter this equation.

- Let  $\mathbf{Z}_i$  be the set of instrumental variables such that the following population moment condition holds:

$$\mathbb{E} [\mathbf{Z}_i \cdot \Delta \xi_{i,j}^s \cdot \mathbf{1}\{t > 0\}] = 0, \forall i, j \quad (61)$$

where  $\mathbf{1}\{t > 0\}$  is an indicator for the periods 1 and 2 for which we take the difference in the neighborhood amenities  $\Delta \xi_{i,j}^s$ .  $\mathbf{Z}_i = [\phi_{h^1, h^2, j}^s, \text{Driving Time}_i]$ , where  $\text{Driving Time}_i$  is the driving time from residential neighborhood  $i$  to the firm's entry site.

---

<sup>57</sup>For brevity in our notations, we omit the subscript  $k$ , which denotes the entry zone. In practice, a pair of locations  $(i, j)$  could be within the 50-mile radius of more than one firm entry site. Further, note that  $\phi_{h^1, h^2, j}^s$  subsumes a constant  $\beta_0^s$  to allow the relative difference in utility between choices inside and outside the entry zone to change across the two periods.

- Define the following moment functions at the  $(i, j)$  level:

$$\psi_1 = \mathbf{Z}_i \Delta \xi_{i,j}^H(\theta) \quad (62)$$

$$\psi_2 = \mathbf{Z}_i \Delta \xi_{i,j}^L(\theta) \quad (63)$$

- We need additional moments from microdata to identify  $b = [b_{in}, b_{out}]$ . We compute model predicted moments to match the observed moments computed using individual's migration history from Infutor. To identify  $b_{in}$ , we match the share of workers born inside the entry zone in period 0, and choose to live in a residential neighborhood  $i$  that is different from his period 0 “birthplace” neighborhood in period 1 and 2, respectively. To identify  $b_{out}$ , we match the share of workers born outside the entry zone but inside the city in period 0, and choose to live in a residential neighborhood  $i$  inside the entry zone in period 1 and 2, respectively. Define the moment functions  $\psi_3, \psi_4, \psi_5, \psi_6$  at the individual incumbent level:

$$\psi_3 = \mathbf{1}\{\omega \text{ lives in } i \neq i_0 \text{ in } t = 1 | i_0 \in \mathcal{N}_1\} - \mu_3, \quad (64)$$

$$\psi_5 = \mathbf{1}\{\omega \text{ lives in } i \in \mathcal{N}_1 \text{ in } t = 1 | i_0 \in \mathcal{N}_0\} - \mu_4, \quad (65)$$

$$\psi_4 = \mathbf{1}\{\omega \text{ lives in } i \neq i_0 \text{ in } t = 2 | i_0 \in \mathcal{N}_1\} - \mu_5, \quad (66)$$

$$\psi_6 = \mathbf{1}\{\omega \text{ lives in } i \in \mathcal{N}_1 \text{ in } t = 2 | i_0 \in \mathcal{N}_0\} - \mu_6, \quad (67)$$

where  $\mu_3, \mu_4, \mu_5, \mu_6$  are the shares observed from Infutor migration data.

- Stack up the these moment functions together:

$$\psi(\theta) = [\psi_1, \psi_2, \psi_3, \psi_4, \psi_5, \psi_6]' \quad (68)$$

- The population moment conditions are assumed to uniquely equal zero at the true parameters  $\theta_0$ , or

$$\mathbb{E}[\psi(\theta_0)] = 0 \quad (69)$$

- Hence we seek estimator  $\hat{\theta}$  that satisfy the sample analogue of the population moment conditions:

$$\hat{\psi} \equiv \frac{1}{N} \sum \psi(\hat{\theta}) = 0, \quad (70)$$

where  $N$  is the number of observations. For the two first two moments, it equals to the number of  $(i, j)$  pairs inside the entry zones. For the last four moments, it equals to the number of incumbent residents in each conditioned “birthplace” group.

- We have four unknown parameters to solve for:  $b_{in}, b_{out}, \beta^H, \beta^L$ .
- The sample analogues of the last four moment conditions are computed as follows:

$$\hat{\psi}_3 = \left( \frac{1}{N_{\mathcal{N}_0}} \sum_{i_0 \in \mathcal{N}_0} N_{i_0}^H \hat{\pi}_{i \neq i_0, t=1}^{H, i_0}(\theta) + N_{i_0}^L \hat{\pi}_{i \neq i_0, t=1}^{L, i_0}(\theta) \right) - \mu_3, \quad (71)$$

$$\hat{\psi}_4 = \left( \frac{1}{N_{\mathcal{N}_1}} \left( N_{i_0}^H \hat{\pi}_{i \in \mathcal{N}_1, t=1}^{H, i_0 \in \mathcal{N}_0}(\theta) + N_{i_0}^L \hat{\pi}_{i \in \mathcal{N}_1, t=1}^{L, i_0 \in \mathcal{N}_0}(\theta) \right) \right) - \mu_4, \quad (72)$$

$$\hat{\psi}_3 = \left( \frac{1}{N_{\mathcal{N}_0}} \sum_{i_0 \in \mathcal{N}_0} N_{i_0}^H \hat{\pi}_{i \neq i_0, t=2}^{H, i_0}(\theta) + N_{i_0}^L \hat{\pi}_{i \neq i_0, t=2}^{L, i_0}(\theta) \right) - \mu_5, \quad (73)$$

$$\hat{\psi}_4 = \left( \frac{1}{N_{\mathcal{N}_1}} \left( N_{i_0}^H \hat{\pi}_{i \in \mathcal{N}_1, t=2}^{H, i_0 \in \mathcal{N}_0}(\theta) + N_{i_0}^L \hat{\pi}_{i \in \mathcal{N}_1, t=2}^{L, i_0 \in \mathcal{N}_0}(\theta) \right) \right) - \mu_6, \quad (74)$$

$$(75)$$

- We follow Hansen (1982), who shows that the optimal (two-step) GMM estimator takes the form

$$\hat{\theta} = \arg \min_{\theta} Q_c(\theta). \quad (76)$$

$$= \arg \min_{\theta} \left( \frac{1}{N} \sum \psi(\hat{\theta}) \right)' C \left( \frac{1}{N} \sum \psi(\hat{\theta}) \right), \quad (77)$$

with a suitable weighting matrix  $C$ .

- The value of the estimate  $\hat{\theta}$  has to be computed using a nonlinear search.
- The optimal weighting matrix  $C$  that minimize  $\text{Var}(\hat{\theta})$  is given by  $\Delta^{-1}$ , the inverse of the asymptotic variance-covariance matrix of the moment functions,  $\Delta = \mathbb{E}[\psi(\theta_0) \psi(\theta_0)']$ .
- Typically we do not know what the variance covariance matrix  $\Delta$  is. Hansen's solution is the following 2-step procedure:
  - \* Estimate  $\theta$  by minimizing  $Q_c(\theta)$  for an arbitrary positive definite symmetric matrix  $C$  to get  $\hat{\theta}_{\text{init}}$ .
  - \* Use this initial estimate  $\hat{\theta}_{\text{init}}$  to get an estimate of the optimal weight matrix ( $6 \times 6$ ):

$$\hat{\Delta}^{-1} = \left[ \frac{1}{N} \sum \psi(\hat{\theta}_{\text{init}}) \psi(\hat{\theta}_{\text{init}})' \right]^{-1}, \quad (78)$$

and use this estimate in the second round by minimizing  $Q_{\hat{\Delta}^{-1}}(\theta)$ .

- Let  $\Gamma = \mathbb{E}[\partial \psi(\theta_0) / \partial \theta']$ , the gradient of the moment functions with respect to the parameters evaluated at the true parameter values. The asymptotic variance of  $\sqrt{N}(\hat{\theta} - \theta_0)$  is given by  $(\Gamma' \Delta^{-1} \Gamma)^{-1}$ .
- we can estimate the asymptotic variance by estimating the two components of the variance,

$\Gamma$  and  $\Delta$ , as

$$\hat{\Gamma} = \frac{1}{N} \sum \frac{\partial \psi(\theta_0)}{\partial \theta'} (\hat{\theta}_{\text{gmm}}), \quad (79)$$

$$\hat{\Delta} = \frac{1}{N} \sum \psi(\hat{\theta}_{\text{gmm}}) \psi(\hat{\theta}_{\text{gmm}})', \quad (80)$$

and substituting these estimates into the expression for the variance.

### Estimating Covariance of Infutor Micro Moments

We use the Delta method to compute the variance-covariance matrix of the Infutor micro-moments. We first define a new set of random variables that yield exactly the same conditional means, but use all observations from Infutor. Define the vector of indicator random variables  $I = [I_1, I_2, I_3, I_4, I_5, I_6]$  for each individual in Infutor as

$$I_1 = \mathbf{1}\{\omega \text{ lives in } i \neq i_0 \text{ in } t = 1\} \times \mathbf{1}\{\omega \text{ born in } i_0 \in \mathcal{N}_1\} \quad (81)$$

$$I_2 = \mathbf{1}\{\omega \text{ lives in } i \neq i_0 \text{ in } t = 1\} \times \mathbf{1}\{\omega \text{ born in } i_0 \in \mathcal{N}_0\} \quad (82)$$

$$I_3 = \mathbf{1}\{\omega \text{ lives in } i \neq i_0 \text{ in } t = 2\} \times \mathbf{1}\{\omega \text{ born in } i_0 \in \mathcal{N}_1\} \quad (83)$$

$$I_4 = \mathbf{1}\{\omega \text{ lives in } i \neq i_0 \text{ in } t = 2\} \times \mathbf{1}\{\omega \text{ born in } i_0 \in \mathcal{N}_0\} \quad (84)$$

$$I_5 = \mathbf{1}\{\omega \text{ born in } i_0 \in \mathcal{N}_1\} \quad (85)$$

$$I_6 = \mathbf{1}\{\omega \text{ born in } i_0 \in \mathcal{N}_0\} \quad (86)$$

Next we construct a function of the means of these six random variables that yields the conditional means of our interest. Define  $h(\cdot)$  such that

$$h(\bar{I}) = \left( \frac{\bar{I}_1}{\bar{I}_5}, \frac{\bar{I}_2}{\bar{I}_6}, \frac{\bar{I}_3}{\bar{I}_5}, \frac{\bar{I}_4}{\bar{I}_6} \right). \quad (87)$$

Note this function gives us the same conditional means as if we were to first select all the individuals that satisfy the condition then taking the average. By the Central Limit Theorem,

$$\sqrt{N}(\bar{I} - \mu_I) \xrightarrow{d} \mathcal{N}(0, \text{Var}(I)), \quad (88)$$

so by the Delta method,

$$\sqrt{N} \begin{pmatrix} \bar{I}_1/\bar{I}_5 & \mathbb{E}[\mathbf{1}\{\omega \text{ lives in } i \neq i_0 \text{ in } t = 1 | i_0 \in \mathcal{N}_1\}] \\ \bar{I}_2/\bar{I}_6 & \mathbb{E}[\mathbf{1}\{\omega \text{ lives in } i \in \mathcal{N}_1 \text{ in } t = 1 | i_0 \in \mathcal{N}_0\}] \\ \bar{I}_3/\bar{I}_5 & \mathbb{E}[\mathbf{1}\{\omega \text{ lives in } i \neq i_0 \text{ in } t = 2 | i_0 \in \mathcal{N}_1\}] \\ \bar{I}_4/\bar{I}_6 & \mathbb{E}[\mathbf{1}\{\omega \text{ lives in } i \in \mathcal{N}_1 \text{ in } t = 2 | i_0 \in \mathcal{N}_0\}] \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \nabla h(\mu_I) \text{Var}(I) \nabla h(\mu_I)'). \quad (89)$$

Hence we can use an estimate of  $\text{Var}(I)$  using the sample of individuals for Infutor and an estimate of  $\nabla h(\mu_I)$  using  $\nabla h(\bar{I})$  to compute the variance-covariance matrix of Infutor micro-moments. The

Jacobian matrix  $\nabla h(\bar{I})$  is:

$$\nabla h(\bar{I}) = \begin{pmatrix} \frac{1}{I_5} & 0 & 0 & 0 & \frac{-\bar{I}_1}{I_5^2} & 0 \\ 0 & \frac{1}{I_6} & 0 & 0 & 0 & \frac{-\bar{I}_2}{I_6^2} \\ 0 & 0 & \frac{1}{I_5} & 0 & \frac{-\bar{I}_3}{I_5^2} & 0 \\ 0 & 0 & 0 & \frac{1}{I_6} & 0 & \frac{-\bar{I}_4}{I_6^2} \end{pmatrix}. \quad (90)$$

### E.9.1 Estimation of Gravity Equation for Commuting Elasticity

From equation of renter's choice probability for pairs of locations  $(i, j)$ , by taking logs of both sides, we can derive the following gravity relationship for commuting times:

$$\begin{aligned} \pi_{i,j,t}^s &= \frac{\exp\left(\frac{w_{j,t}^s - \theta^R r_{i,t} - \lambda^s \tau_{i,j} + a_{i,j,t}^s - b_\omega^s \mathbf{1}_{\omega,t}(i \neq i_0)}{\sigma_s}\right)}{1 + \sum_{i'} \sum_{j'} \exp\left(\frac{w_{j',t}^s - \theta^R r_{i',t} - \lambda^s \tau_{i',j'} + a_{i',j',t}^s - b_\omega^s \mathbf{1}_{\omega,t}(i' \neq i_0)}{\sigma_s}\right)}, \\ \ln(\pi_{i,j,t}^s) &= \frac{w_{j,t}^s - \theta^R r_{i,t} - \lambda^s \tau_{i,j} + a_{i,j,t}^s - b_\omega^s \mathbf{1}_{\omega,t}(i \neq i_0)}{\sigma_s} \\ &\quad - \underbrace{\ln\left(1 + \sum_{i'} \sum_{j'} \exp\left(\frac{w_{j',t}^s - \theta^R r_{i',t} - \lambda^s \tau_{i',j'} + a_{i',j',t}^s - b_\omega^s \mathbf{1}_{\omega,t}(i' \neq i_0)}{\sigma_s}\right)\right)}_{\text{constant}}, \\ \ln(\pi_{i,j,k,t}^s) &= \alpha_i + \rho_j + \eta_k - \left(\frac{\lambda^s}{\sigma_s}\right) \tau_{i,j} + \epsilon_{i,j,t}. \end{aligned}$$

In the last line, we group terms related to  $i$  only into a fixed effect, and group terms related to  $j$  only into a fixed effect. We introduce a fixed effect for entry zone  $k$ . The constant is absorbed into the fixed effects. The error term,  $\epsilon_{i,j,t}$ , summarizes the terms for exogenous amenities and birthplace disamenity, and  $\tau_{i,j}$  and  $\epsilon_{i,j,t}$  are assumed to be uncorrelated. The elasticity of commuting is given by  $\frac{\lambda^s}{\sigma_s}$ . We estimate the regression using choice probabilities in period 1 and 2, and 2009 driving time between pairs of locations  $(i, j)$ , weighted by number of commuters between  $i$  and  $j$ . The driving time is imputed using a similar speed model as we described in C.4, but trained using 2009 NHTS. Table A12 lists the estimated commuting elasticity for period 1 and 2, and for high-skilled and low-skilled workers, respectively. We use 0.02 as the value for commuting elasticity in our welfare decomposition.

### E.10 Change in Expected Welfare for Movers vs. Stayers

We define stayers as incumbent workers who stayed in their period 0 “birthplace” residential neighborhood in both period 1 and period 2; movers are those who stayed in their period 0 “birthplace” residential neighborhood in period 1, and move to a different residential neighborhood in period

2<sup>58</sup>.

In order to calculate the change in expected welfare for movers and stayers, we first calculate the share of workers who belong to the type,  $(i_0, s, o)$ , defined by one's birthplace, skill, and homeownership status prior to the firm entry, among stayers and movers, respectively. For example, by an application of the Bayes' Rule, the share of movers who belong to type  $(i_0, s, o)$  is given by:

$$\begin{aligned} \Pr(i_0, s, o | \text{Mover}) &\equiv \Pr(i_0, s, o | \text{displaced in } t_2, \text{ stayed in } t_1) \\ &= \frac{\Pr(\text{displaced in } t_2, \text{ stayed in } t_1 | i_0, s, o) \times \Pr(i_0, s, o)}{\Pr(\text{displaced in } t_2, \text{ stayed in } t_1)} \\ &= \frac{\Pr(\text{displaced in } t_2 | i_0, s, o) \times \Pr(\text{stayed in } t_1 | i_0, s, o) \times \Pr(i_0, s, o)}{\sum_{i'_0, s', o'} \Pr(\text{displaced in } t_2 | i'_0, s', o') \times \Pr(\text{stayed in } t_1 | i'_0, s', o') \times \Pr(i'_0, s', o')} \end{aligned}$$

For each type  $(s, o, i_0)$ , we calculate the change in expected welfare  $\mathbb{E} [\Delta U^{s, o, i_0}]$ . The change in expected welfare for movers who belong to demographic group  $(s, o)$  is the weighted average of  $\mathbb{E} [\Delta U^{s, o, i_0}]$ , where the weights are  $\Pr(i_0, s, o | \text{Mover})$ :

$$\mathbb{E} [\Delta U^{s, o} | \text{Mover}] = \frac{\sum_{i_0 \in \mathcal{N}} \mathbb{E} [\Delta U^{s, o, i_0}] \times \Pr(i_0, s, o | \text{Mover})}{\sum_{i_0 \in \mathcal{N}} \Pr(i_0, s, o | \text{Mover})}. \quad (91)$$

A similar calculation gives us the change in expected welfare for stayers.

Table A14 compares the welfare incidence for movers vs. stayers. The numbers in brackets are the share of incumbent workers who belong to each demographic group  $(s, o)$  among movers and stayers, respectively. Overall, movers' welfare on average increased by \$12 annually five years after entry, whereas stayers' welfare on average increased by \$4. While the overall difference in welfare changes between the movers and stayers is small, it hides significant heterogeneity in welfare changes among different demographic groups. Among the movers, high-skilled homeowners (27.3% of movers) benefited the most by a \$107 increase in their welfare, while low-skilled renters (22.9% of movers) are hurt the most by a -\$71 decrease in their welfare. Similarly, among stayers, high-skilled homeowners (25.1% of stayers) benefited the most by a \$72 increase in their welfare, while low-skilled renters (22.9% of stayers) are hurt the most by a -\$49 decrease in their welfare. For both movers and stayers, low-skilled homeowners benefited much less and high-skilled renters lose by -\$37 and -\$31, respectively. On net, the welfare effects for different demographic groups cancel out, leading to a small difference between change in welfare between movers vs. stayers. Since there is a higher share of homeowners among the movers, and the homeowners are the real beneficiaries of the firm entries, the overall welfare change is more positive for the movers.

We also notice the difference in the welfare changes between movers and stayers who belong to the same demographic group. Under the set-up of our model, the change in expected utility  $\mathbb{E} [\Delta U^{s, o, i_0}]$  is the same for movers and stayers who belong to the same type  $(s, o, i_0)$ . When they

---

<sup>58</sup>For incumbent workers born outside the entry zone, stayers are those who lived outside the entry zone in both periods; movers are those who lived outside in period 1, and lived inside the entry zone in period 2.

face the same set of options, the decision to move away vs. stay is determined purely based on the differences in the logit error draws in each period. Hence, if movers and stayers have the same initial distribution of birthplace, their  $\mathbb{E}[\Delta U^{s,o}]$  would be the same. Here the difference in  $\mathbb{E}[\Delta U^{s,o}]$  within the same demographic group stems entirely from the difference in initial birthplace distribution between movers and stayers. We know the changes in wages and rents are heterogeneous across neighborhoods, with those closer to the entry site receiving a larger shock. The results that high-skilled homeowners among movers benefited more than those among stayers, and low-skilled renters among movers are hurt more than those among the stayers are consistent with movers are more likely to be initially in neighborhoods that are hit harder by the firm entries.