Startup Location, Local Spillovers and Neighborhood Sorting*

Rolando Campusano†

November 12, 2021

UPDATED FREQUENTLY; CURRENT VERSION AT:

https://rolandocampusano.com/JMP

Abstract

How critical is spatial concentration for the success of startup firms? This paper uses data on the universe of firms in large Canadian cities to study this question at the level of city blocks and their surrounding neighborhoods. To account for sorting within blocks, I use a newly developed clustering algorithm to construct neighborhoods relevant for each industry within which sorting across blocks is conditionally random. To account for sorting across neighborhoods, I develop a model of neighborhood selection, where entrepreneurs choose neighborhoods based on expected startup outcomes and preferences for location. Results show that spillovers of block average same-industry employment and revenue are hyper-local and mostly fade away after 75 meters. These spillovers have economically significant effects on startup end-of-year revenue, employment, and survival rates. For a sense of magnitude, going from the 10th to the 90th industry-specific percentile of incumbent average revenue increases the median startup revenue by 8.2%. These effects are heterogeneous across industries, with employment-intensive industries benefiting relatively more from larger while knowledge-intensive industries benefiting relatively more from better incumbent firms.

*I thank Nathaniel Baum-Snow, William Strange for their excellent mentorship, encouragement, and most of all, their patience with me throughout this long and arduous process. I thank April Franco, Alberto Galasso, and Bernardo Blum for being part of my committee, and their invaluable guidance and support. I also thank Camilo Acosta, Jeremy (Guangbin) Hong, Rallye Shen, Nicolas Grendron-Carrier, Beryl Li, Mark Brown, Ramin Forouzandeh, Pamela Medina-Quispe, Peter Morrow, Victor Aguirregabiria, and participants at the University of Toronto IO and Trade Seminars, the 2021 Canadian Urban Economics Conference and at the 2021 Elevator Pitch of the Urban Economic Association Meetings for many helpful discussions, comments and suggestions. Financial support from the BEAR Ph.D. Award and the TDMDAL Research Grant are gratefully acknowledged. Part of the work was while visiting Finance Canada and Statistics Canada. All errors are my own and do not necessarily represent the views or policies of Statistics Canada.

†Economic Analysis and Policy Area, Rotman School of Management, University of Toronto. E-Mail: Rolando.Campusano@Rotman.Utoronto.Ca. Website: https://rolandocampusano.com.

1 Introduction

This paper is about entrepreneurs, their startups and the spatial scale of productivity spillovers.[1] A substantial amount of evidence for regions and cities, shows that firm and worker productivity are increasing in proximity to higher concentrations of economic activity.[2] Such concentrations occur at many spatial scales, from the city block level up to the regional level. In this paper, I characterized the magnitudes and composition of returns to such proximity at city block and neighborhood level spatial scales. In fact, like Madison Avenue for advertising in New York, Savile Row for tailoring in London, or Bay Street for finance in Toronto, many industries are localized into particular neighborhoods of large cities. In this paper, I estimate the benefits of such localization of incumbent firms to new startup firms.

Productivity spillovers can operate at many different spatial scales. Knowledge flows and labor pooling effects decay rapidly with distance, suggesting that spillovers operating a lower geographical scales might be part of the reasons behind high hyper-local economic concentration (Faggio, Silva, & Strange, 2017; Fujita, Krugman, & Venables, 2001; Rosenthal & Strange, 2001, 2003). These patterns are consistent with the notion that geographic concentration may facilitate the exchange of ideas, promote creativity, and reduce entrepreneurial risk, forces that stimulate the emergence and success of new firms (Alcácer & Chung, 2014; Chatterji, 2009; Chatterji, Glaeser, & Kerr, 2014; Glaeser, Kerr, & Kerr, 2015; Henderson, Kuncoro, & Turner, 1995). The natural implication is a relationship between productivity spillovers that arise with proximity and new firms' success that has motivated countless public and private initiatives that encourage the spatial concentration of new firms (Nathan & Overman, 2013; Porter, 1998). However, there is little causal evidence about the nature of productivity spillovers to new firms, including how they differ by spatial scale.

This paper provides the first causal answer to this question using data on the universe of incorporated firms and their workers in the four largest Canadian cities from 2001 to 2017. While accounting for sorting across industry-specific locations within a city, I find strong evidence of positive productivity spillovers from incumbent firms in the same industry and located in the same block prior the startup operation.[3] I separate these results into scale

[1] It is important to note that along this paper, I refer to entrepreneurs as owners of new firms. This is different from the notion of self-employment that has been also used to characterize entrepreneurs but that omits an important aspect in entrepreneurship, which is making big investments with the hope to make big returns.

[2] A small sample of papers in this broad topic is given by, Arzaghi and Henderson (2008), Combes et al. (2012), Greenstone, Hornbeck, and Moretti (2010), and Rosenthal and Strange (2004).

[3] In terms of identification, I will discuss how estimates in this setting are less likely to be affected by a

effects represented by average size of a company (in number of employees) relative to quality effects represented by average revenue of a company. For startup end-of-year revenue, baseline estimates show an elasticity to average incumbent firms revenue of about 0.02 and to average incumbent firms employment of about 0.12. For startup end-of-year employment instead, baseline estimates show an elasticity to average incumbent firms revenue of about -0.03 and to average incumbent firms employment of about 0.22. For a sense of magnitude, these estimates indicate that going from the 10th to the 90th industry-specific percentile of incumbent firms average revenue increases the median startup revenue by 8.2% and decreases median startup employment by 8.6%. In contrast, going from the 10th to the 90th industry-specific percentile of incumbent firms average employment increases the median startup revenue by 78% and increases median startup employment by 121%.

Consistent with the idea that proximity matters, results show that hyper-proximity maximizes productivity gains for new firms. Compared to estimates in the literature at higher spatial scales, these estimates are relatively large and showcase how important is the geography used to estimate these effects and is believe that there is a modifiable areal unit problem in estimating at larger geographies. This is in part because spillovers operate within blocks, and their effect quickly fades as we move away from the source of exposure. Once including exposure to incumbents one (within 150 meters), two (within 225 meters), or three blocks (within 300 meters) away, the elasticity estimates for exposure within the same block barely change. Moreover, the estimates for the elasticity of the neighboring blocks quickly decay towards zero from one block away in the case of incumbents average revenue and from two blocks away in the case of incumbents average employment.

Different industries are likely to generate different levels and types of productivity spillovers. This is a matter on the mechanisms driving them and how they are operate differently across industries and are likely to affect new firms differently across industries. Differences in results between them showcase that these mechanisms are important and give support, although not causal, to some of these views. In practice, I perform two sets of exercises. First, in the spirit of Faggio, Silva, and Strange (2017), I allow the elasticity to employment and revenue to differ across groups of industries while controlling for differences in market conditions that may arise across them. Second, as in Ellison, Glaeser, and Kerr (2010) I use cross-industry

reflection problem and that the main identification problem in this setting is location sorting. However, as a precaution I use the characteristics of incumbents one year before. Results are similar in both specifications.

connectivity weights to study the importance of the mechanisms underlying these different theories of agglomeration.

Overall, both exercises present a similar picture. Allowing the elasticities to differ across industries shows that employment-intensive industries (such as retail, trade, and manufacturing) benefit relatively more from larger incumbent firms. In contrast, knowledge-intensive industries (such as professional, information, and financial services) benefit relatively more from higher-quality incumbent firms. Results from connectivity weights support the previous results showing that new firms tend to benefit largely from spillovers from more workers in industries with employees performing similar tasks than the startup or higher quality firms in supplier or consumer industries.

Empirically, the main identification challenge in this setting is endogenous location sorting. Firm outcomes are not observed for all locations, but only for the single location in which the startups chooses to start operations. Moreover, as these are new firms, I do not observe their characteristics before choosing a location, which means there can be unobservables also affecting the location choice. In order to overcome this problem, the literature has mainly adopted two approaches depending on the question and data at hand. The first assumes unobservables are constant within central Manhattan, Arzaghi and Henderson (2008), buildings Liu, Rosenthal, and Strange (2018) and commuting zones Roche (2020); or identifying proper location counterfactuals like Greenstone, Hornbeck, and Moretti (2010), Heblich et al. (2019), and Qian and Tan (2021).

I combine both approaches by introducing the distinction between neighborhoods and blocks. Information frictions about the stock of businesses and real estate vacancies introduce uncertainty about the exact industrial characteristic of most blocks the startup may operate. This implies that new firms make location choices based on the expectations about block characteristics. Using rich administrative data on the universe of firms and workers in Canada and additional information about access to amenities, I use machine learnings clustering methods to define neighborhoods in such a way that expectations about all block characteristics within them are statistically identical.

With this distinction I account for location sorting in two steps: across blocks within neighborhoods and across neighborhoods. To account for sorting across blocks within neighborhoods, I construct neighborhoods relevant for each industry within which sorting across blocks is conditionally random. The remaining variation across blocks within neighborhoods in a given year identifies the causal effect of local block characteristics in new firms outcomes. To do this, I use a clustering algorithm developed in a companion paper that shows that neighborhoods are not alike across industries and differ remarkably in size and shape from their administratively defined counterparts (Campusano, 2021). In particular, neighborhoods tend to have an elliptical shape and center around and include both sides of major streets. The algorithm builds upon the idea of propensity score stratification with multiple treatments and incorporates the spatial aspect that characterizes neighborhoods.[4] Effectively, I use historical geocoded firm location choices to estimate a propensity score model using an exhaustive array of location characteristics including industrial composition, land use, amenities and demographic composition. I then group adjacent locations so that resulting location clusters neighborhoods are chosen to minimize the variance of the propensity score within each cluster.

Differences across neighborhoods within and across industries highlight how the same distribution of economic activity affects different industries across a city. The use of neighborhood fixed effects controls for fixed unobserved differences within the boundaries of these newly defined geographies, controlling for sorting into blocks within neighborhoods. However, the interpretation of spillover elasticities in this context is very local as it is using within neighborhoods variation for identification. Using a model to account for sorting across neighborhoods allows me to interpret the elasticities in a more general way, as now they also account for sorting into neighborhoods within a city.

To account for sorting between neighborhoods, I develop a Roy (1951) model of startup neighborhood choices. Identification requires an exclusion restriction that influences entrepreneurs' startup location choices without affecting firm revenue, employment or duration. This poses a challenge for the case of firms as it means assuming that their location decision are not solely made based on their bottom line. However, this is not necessarily the case for startups in this setting as they are controlled by individuals who are most likely to consider their idiosyncratic preferences for location when making location choices for their firm (M. Dahl & Sorenson, 2012; Figueiredo, Guimarães, & Woodward, 2002; Kulchina, 2016).[5] Therefore, in my framework sorting occurs as a consequence of entrepreneur idiosyncratic

[4] A similar approach has been applied by Heblich et al. (2019) and Qian and Tan (2021) for finding suitable sites for new plants. The comparison is not spatial because suitable sites might not be necessarily located next (or even close) to each other. This algorithm extends their approach as it also allows for comparisons across neighborhoods by including their aggregated propensity scores at the neighborhood level.

[5] In fact, a majority of startups in my sample are founded in the place of residence of one of the founders.

preferences.

To estimate this model, I use control function methods developed by G. Dahl (2002) and Lee (1983). In particular, controlling for a nonparametric function of location choice probabilities interacted with neighborhood fixed effects accounts for non-random sorting into neighborhoods.[6] However, estimating the choice probabilities in this setting is challenging. The identification strategy across neighborhoods requires having choice probabilities at the entrepreneur-neighborhood level. All of this implies that I require probabilities at the entrepreneur-neighborhood-year level.

Using standard methods to estimates the choice probabilities is infeasible in a setting with that many alternatives. To resolve this issue, I use Kingma and Ba (2017) Adam's algorithm to estimate a neural network in which each entrepreneur has a unique node in the network. Each node provides information about the relationship between features and choices and informs the rest of the network through forward and backward loops. Outputs of each node are propagated (and back-propagated) through the network, exploiting the ability of early layers to learn representations that can be utilized by later layers in the network.

None of this would have been possible without the Canadian administrative tax data. Like many other matched data sets, this data set includes detailed balance sheet information for corporations and income tax information for workers. However, this data has three additional features that are crucial for this analysis. For every firm and individual, I know its location at the block-face level. Having such detailed information is crucial as it facilitates not only the characterization of the industrial environment at fine geographies, but it also allows me to construct industry-specific relevant neighborhoods, which is crucial for identification. For every firm I also know every owner (corporation or individual) with more than 10% of the shares of a corporation, and for every firm that I know if a large proportion of its employment used to belong to a firm that is not in the data anymore. I use this information for two important reasons. First, I identify true new firms from other type of firms that appear to be new in the data but they are not, such as new branches/subsidiaries or merge and acquisitions. Second, it is crucial for identification of the model that corrects for sorting across neighborhoods, as it allows me to include owners characteristics.

This paper mainly contributes to the intersection of two literatures: the literature in

[6] This polynomial serves as a multidimensional analog of the inverse Mills ratio in the classic Heckman (1979) model.

urban economics that studies the effects and incidence of spatial treatments on firms, and the literature in entrepreneurship that seeks to understand how the local environment affects new firms success. Besides the uniqueness of this data for this context, the main contribution is done by recognizing the identification challenges that this type of question poses and incorporating tools from the literature of treatment effects and computer science to address them. Models of self-selection like the one used in this paper have long been used to understand individual treatment choices and its effect over their own outcomes. Recognizing that startups are owned by individuals who not only consider their firms but also their own preferences, and having data about that, makes using this type of model possible. Moreover, recognizing that neighborhoods are collections of similar micro-locations justifies use of a propensity score method that applies machine learning clustering algorithms. Finally, treating neighborhood choice as a classification problem facilitates the use of neural networks to estimate probabilities that are otherwise impossible to estimate with current methods.[7]

The rest of the paper is organized as follows: Section 2 describes the data, sample and measurement choices, Section 3 discusses the main identification challenge and introduces the distinction between neighborhoods and blocks. Section 4 provides the main results for block level spillovers within neighborhoods. Section 5 describes the neighborhood choice model, the empirical methodology to control for sorting across neighborhoods, and provides the main results for block level spillovers across neighborhoods, and Section 6 concludes.

2 Data

The primary data source is the 2019 vintage of the Canadian Employer-Employee Dynamics Database (CEEDD) produced by Statistics Canada. The CEEDD links records of employment (T4 Statement of Remuneration Paid) to individuals (T1 Individual Income Tax and Benefit Returns) and corporations (T2 Corporation Income Tax Returns) for the universe of workers, firms and owners in Canada. The T4 files contain information on individual-firm specific income and taxes paid. The T1 files contain information on individual characteristics, family composition and income. The T2 files contain information on firm revenue, expenses, and assets. Additional information on payroll and employment is derived from the linked T2

[7] This is related to the observation made by Mullainathan and Spiess (2017) and Athey and Imbens (2019) that machine learning not only provides new tools, but also solves different problems than current empirical methods.

and T4 records.

I keep all firm-years in the Calgary, Montreal, Toronto and Vancouver CMAs with some evidence that the firm is operating.[8] I focus on using information about sales of goods and services (revenue), employment, and payroll as these are required reporting lines in the corporate tax filings. I drop firms that cycle back and forth between postal codes, with missing location information, or with no 4-digit industry information. I identify a firm's founding and closing years as the first and last years it has positive reported revenue, employment and/or payroll. Multi-locations firms only report one location per firm, which I take to be their headquarters.[9] I new drop multi-location firms, and I keep all incumbent firms as they are likely to produce important spillovers for new firms.[10] I drop firms in NAICS s,2, 6 and 9 as their location choices are generally constrained by geographical features (in the case of Agriculture and Mining) and politically defined boundaries (in the case of Health, Education and Public Administration).

The Canadian data has two crucial features that differentiates it from other sources that contain a similar employer-employee match structure. For every firm I know its almost-complete ownership structure, and for every firm and individual I know its location at the block-face level. Individuals and firms locations are obtained as Canadians need to report their address of residence (in the case of individuals), and their address of operation and legal activities (in the case of corporations). The ownership structure is obtained as Canadian corporations must declare every corporation or individual that owns more than 10% of shares (T2 Schedule 50).[11]

Even though firms and individuals report full addresses, only six-digit postal codes are released. However, Canadian postal codes in major cities are very small as they typically cover a single side of a city block (a block-face), a single building or even a portion of a

[8] These cities correspond to the largest 4 Canadian cities. Other major CMAs were left out mainly because of the size of their postal codes (Edmonton) or city specific particularities. For example, Ottawa-Gatineau is a mainly dominated by government establishments and workers.

[9] This implies that big corporations with many establishments will produce a large concentration of labor and revenues in their headquarter locations.

[10] In this case we can think of total employment or total revenues as proxies of the importance of the firm more than the level in a given location.

[11] The forms contain much more information that what it's used in this paper. For reference about what is available in them please visit:

T1: ww.canada.ca/en/revenue-agency/services/forms-publications/tax-packages-years/general-income-tax-benefit-package.html

T4: www.canada.ca/en/revenue-agency/services/forms-publications/forms/t4.html

T2: www.canada.ca/en/revenue-agency/services/forms-publications/forms/t2.html

T2SCH50: www.canada.ca/en/revenue-agency/services/forms-publications/forms/t2sch50.html large building. This means that using the coordinates of the centroid of each postal code provides a precise estimate of the location of firms and individuals in major cities. Having such detailed information is crucial as it allows not only the identification of firm neighbors and their characteristics but also the use the spatial structure of the data for identification purposes. This is done by constructing local counterfactual locations for firms in different industries, which introduces the distinction between neighborhoods from blocks that will be used throughout the rest of the paper (I defer this discussion to the next section).

The ownership structure provides the identifiers of firms and individuals than owns more than 10% of a firm shares in a given year. Linking these identifiers back to T1, T2 and T4 files allows me to obtain among other things owners characteristics (both individuals and firms), and their job and location history. Moreover, it allows me to separate new firms mainly created by individuals or a group of individuals from new firms mainly created by other firms. This distinction is important as this paper focuses on the former group of firms that are more related to the model of entrepreneurship presented below.

2.1 Firm Classification

This paper focuses on how nearby incumbent firms affects new firms. However, separately identifying them is challenging (Decker et al., 2014; Haltiwanger, Jarmin, & Miranda, 2012; Schoar, 2010). Most administrative tax data do not have a specific flag for "new firms", which means that they can only be isolated by new entries in the dataset in a given year. However, as firms in these data are defined based on tax reporting units; mergers and acquisitions might appear in the dataset as a new entry while in fact an incumbent firm is simply changing tax reporting units. For example, if a large firm acquires a small firm, the small firm seems to have died, and the consolidated firm might appear in the data with its own new identifier. In reality, the acquisition resulted in neither a firm birth nor a death. The employees of the large firm still work for the same large firm, and most employees of the small firm work for the same establishments, and hence for the new larger firm but under a new ownership.

To deal with this challenge I use a series of links between firms developed by the Longitudinal Employment Analysis Program at Statistics Canada (J. R. Baldwin, Landry, & Leung, 2016; Dixon & Rollin, 2012). These links identify groups of firms that open or close one year apart and are potentially related but have different tax identifiers. In particular, a pair of firms is flagged as connected if they share a substantial portion of employees, belong to the same industrial classification and share similar firm name and addresses.[12] The result is observation of many different types of linkages between firms, depending on the complexity of the restructuration the firms are undertaking. The key if the ability to separately observe incumbents from true entrants.

After identifying truly new firms, the next challenge is to separate those new firms that are created from existing firms from those started by individuals with little-to-no business ownership experience. The second group are more related to the notion of entrepreneurship examined here.[13] These firms are called *de novo* firms in the management literature. It is important to note that even though these firms lack of a parent company, they might have corporate partners. Moreover, these type of firms also include spin-outs which are new firms started by employees from incumbent firms (Franco & Filson, 2006).

Using the ownership structure of each firm, I to identify both types of new firms. The database reports every owner with at least 10% of common or preferred shares. I use this structure to classify new firms into groups: non-institutional and institutional firms. Institutional firms are those that either have a parent company that owns a majority stake of common shares or have their ownership diluted in many owners that own less than 10% of common shares, which would be the case for most publicly owned firms. Non-institutional firms are the complement. Firms owned by one or few individuals that have a majority stake in the company's common shares. Within these companies, I define the controller as the owner that owns the majority of the company, the owner that owns the majority within the family that owns the majority of the company and/or the owner that owns the maximum share of the company and also works for the company (has a T4 associated with the company). Finally, I flag the existence of preferred shares as well as a minority corporate partner as a signal of having investors supporting the company.

Firm characterization

Altogether, there are three types of firms; startups, institutional entrants and incumbents. Table 1 presents summary statistics for single-establishment firms operating in at least one year between 2001 to 2017 in Calgary, Montreal, Toronto, and Vancouver. Overall there are 1,347,920 firms and 8,439,646 firm-years. Out of all these firms, 565,970 started operations

[12] See J. Baldwin, Dupuy, and Penner (1992) for a description of the construction of the database.

[13] The first group of firms are called *de alio* firms in the management literature or spin-off firms in the finance literature. I interpret both concepts to be very similar (Agarwal & Moeen, 2016) .

from 2002, 317,606 of them were startups, and 284,364 were institutional entrants. Each column show statistics for incumbent firms, startups, and institutional entrants. The most important takeaway of this table is that startups and institutional entrants and nothing alike. The typical institutional entrant has end of year revenue 10 times higher than the typical startup. Moreover, it has employment that is almost 5 times higher than the typical startup. These startups are sufficiently small that entering to a given block is unlikely to influence incumbents and local environment.

[Table 1 about here.]

2.2 Location characteristics

In large cities, Canadian postal codes denotes a single side of a city block, a single large building or a portion of a very large one, or a business that receives large volumes of mail on a regular basis. Moreover, postal codes tend to change over time in size and form to adjust to the needs of mailing operations. For these reasons using postal codes to characterize locations may lead to measurement error. To overcome this issue, I assign postal codes to hexagons of 75 meter sides. Therefore, the smallest geographical unit in this paper is given by blocks, which from now onwards are equivalent to postal codes or many postal codes whose centroid is located inside the same hexagon. This means that location characteristics are defined at the hexagon-cell level. Figure (1) shows how hexagons are assigned and the rings around them. I will use these rings to study spatial decay in the estimation section.

[Figure 1 about here.]

[Table 2 about here.]

Table (2) report statistics about firm locations. As in Table (1), each column reports means of block-level characteristics for the three types of firms. The first row indicates that a very large proportion of firms operates in the place of residence of the main controller.[14] Moreover, it shows that there is also a high proportion of new firms that move within five years of starting operations. This is consistent with a difficult decision, which I will use later for identification

[14] In this case of startups this is given by one of the founders. However, in the case of Inst. Entrants it is given by the place of operations of the parent company, and in the case of incumbents is given by either the residence of individual owners or place of operations of the parent company.

purposes. Finally, this table also shows that startups start operations in places with less concentration of economic activity, with firms with lower revenue and employment relative to the other types of firms, but at the same time in places with high dynamism and new firms.

From now on, I will focus on the effect of incumbent firms on startups. Unless stated otherwise, I do not consider institutional entrants in any of the following analysis. The next section will introduce the main identification problem and how I address it using a combination of machine learning methods and economic theory.

3 Economic Neighborhoods

As discussed in the Introduction, a key element in the research design is the construction of neighborhoods. In contrast to typical administrative neighborhoods, these neighborhoods are constructed taking into account firms objectives and the way they differ across industries. Consider all new firms who have already made their first location decisions and begun operating. The population end-of-first-year outcome function $y_{lt}^{(i,j)}$ for new-firm i of industry j that began operations at time t in location l is given by:

$$y_{lt}^{(i,j)} = x_t^{(i,j)}\beta + X_{lt}^{(j)}\theta + \lambda_t^{(j)} + \epsilon_{lt}^{(i,j)}. \tag{1}$$

Firm i's end of year outcome $y_{lt}^{(i,j)}$ depends on its own observed characteristics $x_t^{(i,j)}$, an aggregation of characteristics of incumbent firms of industry j in location l at time t, $X_{lt}^{(j)}$, industry and time specific fixed effects, $\lambda_t^{(j)}$, and a transitory shock $\epsilon_{lt}^{(i,j)}$. This specification implies that firm's outcome $y_{lt}^{(i,j)}$ may be influenced, not only by its own characteristics, but also by the characteristics of incumbent firms who interact with it. The important point here is that firm interactions arise via spatial proximity.

These interactions produce spillovers across firms that are captured by θ. The primary threat to identification is that firm's outcomes are not observed for all locations, but only for the single location in which the firm chooses to operate. This means that, in the self-selected sample for location l, the error term $\epsilon_{lt}^{(i,j)}$ does not necessarily have zero mean conditional on the independent variables, and ordinary least squares regression potentially yields biased estimates of β, θ and $\lambda_t^{(j)}$. The ideal experiment would be to randomly assign twin firms to different locations and then compare firm outcomes. As an experiment like this is impossible, the literature has adopted two different approaches depending on the question and data at hand. The first is to assume unobservables are constant within a certain broad geographic area as in Arzaghi and Henderson (2008), Liu, Rosenthal, and Strange (2017), and Roche (2020) for Manhattan, buildings and commuting zones respectively. The second is to use difference in difference strategies as in Greenstone, Hornbeck, and Moretti (2010), Heblich et al. (2019), and Qian and Tan (2021).

I built on both approaches. Assuming location unobservables are constant within a certain broad geographic area facilitates the use of broad area-time fixed effects to identify block level spillovers. The identifying variation in this case comes from different outcomes of all new firms that start operations the same year in the broad area but in different blocks within that area. If location unobservables propagate in space, then the validity of this assumption decreases as the size of these areas increases, as it is more likely that blocks unobservables differ within broad areas. The challenge is to identify areas that are large enough to have block-level variation in startup outcomes and small enough such that block location is conditionally random.

The key to my approach is the distinction between neighborhoods and blocks.[15] Informational frictions about the current stock of business sites and vacancies introduce uncertainty about the exact industrial characteristic of a block. This implies that new firms make location choices based on expectations at the block level. Those expectations propagate spatially creating neighborhoods composed by adjacent blocks that are otherwise equal in the eyes of entrepreneurs. New firms make location choices at the neighborhood level and are then subject to vacancies arising randomly within that area. Such neighborhoods are different across industries.[16]

3.1 Constructing Economic Neighborhoods

Neighborhoods are economic in nature as they result from the collective choices of different entities. Startup owners consider the characteristics of economic neighborhoods when making

[15] In terms of notation, this means redefining the subscript that identifies a location from l to bn that represents a block b located in neighborhood n.

[16] For example, restaurants, hotels and touristic attractions might care differently about location characteristics such as access to foot traffic, financial services or transportation infrastructure, than firms in the manufacturing or the professional services sector.

location choices. As these firms are small and face informational constraints, they take those characteristics as given. As in Arzaghi and Henderson (2008), Liu, Rosenthal, and Strange (2017), and Roche (2020), variation within a broad area (neighborhoods) provides identification of block-level spillovers within that area. As in Greenstone, Hornbeck, and Moretti (2010), counterfactuals locations (blocks within the neighborhood) are different across type of firms (industries) reconciling the fact that different firms value location characteristics differently.

I measure block attributes using the universe of firms in the four major Canadian cities for the years 2002 to 2006. I estimate a block level propensity score each industry and city. The propensity score parsimoniously captures the likelihood that a block is suitable for a firm in a given industry.[17] Then, neighborhoods are constructed following the ideas of propensity score matching and propensity score stratification, in which instances that belong to the same bin of propensity score are comparable. In particular, using the spatial clustering algorithm developed in Campusano (2021), neighborhoods group adjacent blocks that have propensity scores that are statistically indistinguishable from each other.

This and the next subsection summarizes the key aspects of the computation of the propensity score and how it's used to identify neighborhoods. Hereafter I use 'economic neighborhoods' to distinguish neighborhoods that result from the clustering of blocks, from other definitions of neighborhood boundaries such as postal codes, counties or census tracts.

3.1.1 Propensity Score

Neighborhoods are constructed under the presumption that a pre-choice propensity score well represents the desirability of a block for for the entry of a new firm in a given industry. Therefore, if two adjacent blocks have an indistinguishable difference in the score means that entrepreneurs are ex-ante indifferent between the two locations and the blocks are treated as one unit. However, given that block characteristics vary over time, they are still able to causally affect new firms outcomes, provided two conditions are satisfied. First, the propensity score must be a sufficient statistic of the expected desirability of a block. Second, entrepreneurs have imperfect information about the future evolution of a block. In other words, within neighborhoods block-level economic conditions do not affect the probability a

[17]Heblich et al. (2019) and Qian and Tan (2021) use a similar approach for finding counterfactual location choices for new plants.

block receives a new firm.

The first condition is satisfied by using a comprehensive and detailed set of block-level characteristics that drive firm location choices. These will be constructed for a pre-period to address reflection issues. When a new firm chooses a location, its desirability is influenced by supply and demand factors that matter differently across industries and cities. For these reasons, I use data for the universe of firms (incumbents and startups) and workers to construct a database of block-level characteristics before new firms make location choices. I obtain amenities data at the block level using a rich point of interest dataset. Finally, I compute the block level propensity score for each city and industry.

For each block, I compute the following measures of access to amenities, workers, firms and residents for the period between 2002 and 2007.[18] I use CEEDD to compute the number of firms operating in each industry and the number of workers that reside and or work there. I use the weights in the 2002 Statistics Canada's input-output matrix to compute the number of firms and workers in industries upstream and downstream the production process. I then use the DMTI Spatial Inc. CanMap Postal Codes Enhanced Suite to compute access to amenities. In particular, I compute the percentage of the land of a block that is used for each of the following categories: Commercial, Government, Industrial, Parks, or Residential. Finally, I compute the number of establishments in each of the following categories: Hotels, Schools, Banks, Hospitals, Touristic Attractions, and Police and Fire Stations.[19]

Finally, for each block and industry, I predict the average annual probability of receiving a firm in their industry using a conditional logit regression in which observed location choices reveal preferences associated with the firm's profit maximization problem (McFadden, 1973). In practice, applying this approach to any location choice poses a problem related to the definition of the spatial choice set. Ideally, small areas should be used, because factors usually identified as relevant for location decisions apply at the local level and consequently cannot be adequately taken into account when the model considers large areas in the spatial choice set.

[18] The benefits (or costs) associated with a block are transmitted spatially. This means that the distribution of these variables is smoother over the space. I incorporate this concept by calculating each block-level access measure as the exponentially discounted-by-distance sum of each measure between a block and immediately adjacent blocks.

[19] According to DMTI, this dataset comprises the location of the universe of Canadian business and recreational points of interest at the establishment level. Among other things, it also includes land use information at the block level.

However, the use of small areas poses a difficulty for estimation, as the conditional logit model does not handle large choice sets very well. Guimarães, Figueirdo, and Woodward (2003) provides a solution to this problem by demonstrating that, under the assumption that individual decisions are based exclusively on a vector of choice-specific attribute variables common to all decision-makers, the coefficients of the conditional logit model can be equivalently estimated using a Poisson regression, which I use instead.[20] Appendix **??** provides more details.

3.1.2 Neighborhoods as propensity score clusters

By using all firms in a given industry, the propensity score summarizes how an industry internalizes the effect of different location characteristics in profits. By predicting the probabilities for different industries, the differences in propensity score across industries showcase how industries value location characteristics differently. To illustrate this, Figure 2 shows the spatial distribution of the predicted probability for the manufacturing and entertainment industries in a one kilometer radius around a major intersection in the city of Toronto.[21] As expected, there is important variation across industry in the propensity score. Certain locations within a city that might be desirable for one industry, might not be desirable for another. Moreover, the way the propensity score propagates in space differs across industries. The standard approach to neighborhoods is to use administrative or legally-defined boundaries. However, this can be a problem as administrative neighborhoods are delineated as a result of an optimization process that does not necessarily align with agents' decision problems. For example, postal codes are defined for the purpose of optimizing mail delivery and census blocks (or tracts) are defined for the purpose of optimizing the process of a census. A persistent misalignment between 'legal' and 'economic' boundaries that may lead to measurement and inference biases that compromise research findings and the policies designed around them. Figure 3 superpose administrative boundaries to panel (a) of Figure 2 and shows a clear spatial misalignment.

For these reasons, I use Campusano (2021) proposed revealed preference approach to delineate economic neighborhood boundaries. The algorithm uses historical geocoded location choices to identify neighborhoods as a collection of *similar-neighboring-choices.* In particular, neighborhoods are constructed by grouping together contiguous blocks based on the probability of observing a new firm in a given block. This is very similar in spirit to Rosenbaum

[20] Table (**??**) contains the complete list of characteristics in the model. Table (**??**) shows the estimated parameters of the propensity score model. **Discussion of results.**

[21] To maintain confidentiality of CEED data, the estimated probabilities in these maps are the result of using only the external DMTI data.

and Rubin (1984)'s propensity score stratification that has been proved to significantly reduce the bias due to measured confounders when estimating linear treatment effects. Within each propensity score stratum, treated and untreated subjects have roughly similar values of the propensity score. Therefore, when the propensity score has been correctly specified, the distribution of measured baseline covariates is approximately similar between treated and untreated subjects within the same stratum.

[Figure 2 about here.]

[Figure 3 about here.]

Strata are formed by defining a threshold by which the data is divided in a previously set number of stratum. This is done by ordering the estimated propensity score for each location and then grouping them. This has some drawbacks. First, one needs to define either the threshold or the number of stratum beforehand which can lead to some degree of bias although it has been proved that by stratifying in something greater than quintiles there is a substantial reduction in confounding bias.[22] Second and most importantly for this paper, defining stratum this way implies neighborhoods that are not compact in the space. Campusano (2021) addresses these drawbacks by using semi-supervised clustering methods.[23][24]

In practice the algorithm works as follow. After computing the propensity score at the block level, I compute a dissimilarity matrix by calculating the bilateral euclidean distance only between adjacent blocks. This creates a sparse matrix populated by non-zero elements around the diagonal. Adjacency constraints are enforced by assigning an arbitrarily large number for the rest of the elements. The algorithm starts with each block as a neighborhood by itself. Further iterations merge one neighborhood at a time so that the bilateral distance between the neighborhood and the new elements are below a given threshold and together

[22]In fact, increasing the number of strata used should result in improved bias reduction, although the marginal reduction in bias decreases as the number of strata increases (Cochran, 1968).

[23]It is a semi-supervised algorithm since the algorithm incorporates economic theory. Adding adjacency constraints means I have an economic prior on how this clusters should look like, and by clustering over an estimated propensity score I am assuming that the weights assigned to different attributes comes from a location choice model.

[24]For more details about the properties of the clustering method, please refer to Campusano (2021). For more details about this specific application please refer to Appendix (**??**).

minimize the within neighborhood variance in the propensity score.[25] As an illustration, Figure 4 showcases the same location as in the previous figures but now it overlays some of the agglomerating steps undertaken in order to get to a neighborhood.

[Figure 4 about here.]

3.2 Characteristics of Economic Neighborhoods

I apply the algorithm to the estimated propensity score for each industry in each city. I use five different thresholds and define them as functions of the standard deviation of the propensity score. The thresholds are set at $2, 1, 0.5, 0.1$ and 0.01 standard deviations of the propensity score. Neighborhoods are composed of at least two grid-cells. Identification of block-level spillovers relies on the assumption that, within neighborhoods blocks cannot be distinguished. Put differently, it relies on the assumption that sorting occurs at the neighborhood level only, meaning that startups locations within neighborhoods cannot be predicted using block contemporaneous characteristics. I choose the threshold taking this into account by selecting, for every industry and city, the maximum threshold that delivers neighborhoods in which the estimated propensity score is uncorrelated or minimizes the correlation with the number of new firms in a block within a neighborhood.

I test the resulting neighborhoods in the estimation sample (2002 to 2017) by computing a propensity score for every block in every city and industry. I then correlate the contemporaneous propensity score with the contemporaneous number of new firms in a block controlling for neighborhood-year fixed effects. Table (3) reports the results of this exercise, and shows that within neighborhoods for most cities and industries, the contemporaneous propensity score is uncorrelated with the contemporaneous number of new firms. This provides a good test for the degree to which controlling for neighborhood-years accounts for sorting, and it shows that for some industries and cities the method provides better neighborhoods than for others.[26]

[25] The use of thresholds is unavoidable for any approach that seeks to discretize a continuous territory (de Bellefon et al., 2019). However, Campusano (2021) shows that in situations in which the distribution of economic activity is highly skewed, neighborhoods are stable to changes in the threshold decreasing the concerns about the use of an previously defined threshold. In any case, robustness to the threshold are provided in the appendix.

[26] The application of the algorithm for Montreal seems to deliver neighborhoods that are not as good as the neighborhoods for other cities. One possible explanation that requires further investigation is the remarkable differences in legal system and culture that is between Montreal and the other cities in my

Figure 5 serves as an example of the resulting neighborhoods. To maintain confidentiality of CEED data, these maps are the result of using only the external DMTI data. The neighborhoods obtained using the full data are consistent with these maps and provide a better approximation to economic neighborhoods. The main result from the exercise is that neighborhoods are different from each other across industries. Typically industries that have less reach due to window shopping or less within city tradability, such as professional services and retail, have smaller neighborhoods than industries that are more tradable within a city, like manufacturing and entertainment services. In general, neighborhoods can be represented by a shape longer than it is wide, like a rectangle or an ellipse. Further visual inspection shows that these neighborhoods tend to locate around major streets. This is true for all industries, but more so for the manufacturing and trade industries. Moreover, the smaller area of professional services and retail is given more by differences in length than in width, with neighborhoods that are between 42% to 59% shorter than the manufacturing and entertainment services.

[Table 3 about here.]

[Figure 5 about here.]

4 Local Spillovers

This section employs the economic neighborhoods constructed above to identify spillovers at a hyper-local setting within neighborhoods. In this research design, adding neighborhood-year fixed effects solves for the endogeneity in the error term of equation (1) and delivers identification of the causal effects of block level characteristics over startup outcomes. The new estimating equation for firm i's starting operations in block b of neighborhood n at time t turns into

$$y_{bnt}^{(i,j)} = x_t^{(i,j)}\beta + X_{bnt}^{(j)}\theta + \lambda_t^{(j)} + \delta_{nt}^{(j)} + \epsilon_{bnt}^{(i,j)} \qquad (2)$$

where the end-of-year outcome $y_{bnt}^{(i,j)}$ depends on her own observed characteristics $x_t^{(i,j)}$, an aggregation of characteristics of incumbent firms of industry j in block b within neighborhood

sample. This implies that location choices in Montreal are different than in the other cities and that the propensity score is failing to account for them. This has implications over the algorithm.

n at time t, $X_{bnt}^{(j)}$, industry and time specific fixed effects, $\lambda_t^{(j)}$, neighborhood-year fixed effects, $\delta_{nt}^{(j)}$, and a transitory shock $\epsilon_{lt}^{(i,j)}$.

Table (2) in section (2) shows that a large fraction of new firms decide to start operations in the main owners's place of residence. In that scenario, it is most likely that the entrepreneur has previous knowledge about the industrial environment of her place of residence. For that reason, in this section I only consider new firms that open away from the entrepreneurs' place of residence. This restriction will be relaxed in the next section where I model the neighborhood choice decision.

4.1 Main Results

The coefficient of interest θ represent effect of the level and quality of incumbent firms over startup outcomes. Provided that industrial, city and time differences are accounted for, higher levels of incumbent employment and revenue are good measures of better exposure to large and high quality firms. In particular, incumbent characteristics are measured as the average employment and revenue of the firms in the same block one year prior the entry of the startup. In this specification, identification of the elasticities is coming from variation in the outcomes of otherwise similar startups randomly assigned to distinct but otherwise similar blocks within a neighborhood. This means that controlling for the number of incumbent firms, I interpret the elasticities of startups' outcomes to these characteristics as scale (average employment) relative to quality (average revenue) spillovers.

In addition to the main predictors and set of fixed effects, I include a series of controls and firm characteristics that aim to account for "prior-to-launch" differences across startups and for potential competition effects that might arise due to higher concentration of economic activity. Prior-to-launch differences are a threat to identification as they might be correlated with how "subject to vacancies" these new firms are. For example, entrepreneurs with previous industry experience might have better knowledge about the differences across blocks within a neighborhood, or firms with corporate partners might be able to delay entry if the vacancy that arrives is not optimal. I control for these and other potential threats by including the owners' years of experience in the industry, a dummy variable for startups with corporate partners, a dummy variable for startups that are controlled by many members of the same family, and the number of prior businesses owned by the entrepreneurs. Finally, I also include the number of incumbent firms in the block and a dummy variable when the the block has no firms or their firms have either zero revenue or employment.

[Table 4 about here.]

New firms often have either zero employment or revenue, meaning that estimating equation (2) in logs is not feasible. Recent research in the international trade literature makes the case for using Poisson pseudo maximum likelihood (PPML) in scenarios with highly granular data. As such, I implement PPML with high dimensional fixed effects (Bailey et al., 2021; Silva & Tenreyro, 2006).[27] All standard errors are clustered at the neighborhood-year level.

Table (4) shows the main results of the paper using end-of-year revenue (in panel a) and end-of-year employment (in panel b) as outcomes of interest. Column (1) shows the preferred specification for firms choosing locations away from the owners' places of residence using neighborhood-year fixed effects, the set of controls mentioned previously and only neighborhoods that exhibit at least two new firms starting operations in different blocks. The remaining columns show a series of exercises that test for robustness and identification of the main results to different samples and specifications.

I find strong evidence of positive productivity spillovers of incumbents' of the same industry and located in the same block prior the firm starts operating. Within industry-specific neighborhoods, exposure to larger incumbents increases both startups end-of-year revenue and employment. While exposure to better incumbents increases startup end-of-year revenue, it decreases startup end-of-year employment. I interpret these spillovers as productivity spillovers increasing startups performance through outputs and efficiency. Column (1) shows that, for end-of-year revenue, the elasticity to average incumbent revenue is about 0.02 and the elasticity to average incumbent employment is about 0.12. For end-of-year employment instead, estimates show an elasticity to average incumbent revenue of about -0.03 and to average incumbent employment of about 0.22.

To give a sense of magnitude, these estimates indicate that going from the 10th to the 90th industry-specific percentile of incumbent average employment increases the median startup revenue by 78% and increases median startup employment by 121%. Moreover, they indicate that going from the 10th to the 90th industry-specific percentile of incumbent average revenue increases the median startup revenue by 8.2% and decreases median startup employment by

[27] Estimation of these type of models are subject to a problem of statistical separation. This problem arises when fixed effects are able to separately identify observations without providing extra information for the estimation of the parameters. This leads to convergence problems and/or incorrect estimates that are corrected using Correia, Guimarães, and Zylkin (2020) estimation.

8.6%. This means that accounting for industry differences, higher revenue incumbents affects outcomes of new firms through helping new firms to achieve higher outputs and efficiency, while larger incumbent firms affects outcomes solely through helping new firms to scale up. These numbers are large and represent the importance of exposure to concentration of economic activity for new firms.

The identification of these elasticities rely on fixed effects accounting for sorting across neighborhoods in a given year. In that case, a threat to identification is the use of neighborhoods constructed using the same data source as the one used for estimation. As mentioned in the previous section, I address that threat by using a combination of a fixed pre-period of the data with external data to only construct the neighborhoods and not for the estimation. However, even though path dependency is unlikely in a growing cities during a period of more than 10 years, it is still possible and could potentially bias the estimates. Columns (2) and (3) test for this by not using neighborhood-year fixed effects or using 3-digits postal codes instead. Results of these exercises provide larger estimates of the elasticities which is consistent with upward biased estimates due location sorting, and with neighborhood-year fixed effects helping to account for that. Column (4) test for differences between otherwise similar firms that might affect sorting within neighborhoods. Column (5), (6) and (7) test for sample definitions. Column (8) and (9) test for estimation method. All tests lead to the same conclusion of positive local spillovers of incumbents over new firms.

Dynamics Location choices are dynamic. This is embedded in the observed sorting patterns in the data, which are accounted for with neighborhood fixed effects. The natural implication is to analyze the degree to which spillovers that are positive in the short term persist over time. Table (5) reports the effects of location attributes over firm survival and future moving choices. These estimates needs to be taken with caution as the larger the time span after the founding year the more likely is the startup to affect incumbent firms, and hence estimates would be biased. In that sense is more likely that estimates one year ahead are better identified than estimates five year ahead.

Columns (1) and (2) report estimates from a specification identical to column (1) of Table (4), except that the outcome of interest is given by an indicator variable equal to one if the firm is alive one year after starting operations (column 1), if it is alive five years after start operations (column 2). Both columns portrait a picture consistent with the main results, spillovers are positive and they materialize in higher probability of survival for startups exposed to higher economic activity. However, these estimates show that exposure to larger firms is better for firm survival than exposure to firms with higher revenues. This is consistent with a story that implies startups first growing in size before growing in sales.

[Table 5 about here.]

Firms may wish to move across cities as they evolve. Columns (3) to (6) report the extend to which early exposure to higher concentration of economic activity affects future location choices. Results show that exposure to larger firms in the same industry decreases the likelihood of moving one and five years after founding the company. However, exposure to firms with larger revenue increases the likelihood of moving five years after founding the company. There are many reasons why this would be. For example, firms may wish to accrue knowledge gains in other location, like workers moving from large cities and taking a portion of their urban wage premium with them (Glaeser & Maré, 2001). Firms may also wish to move if their needs change with growth requiring different types of inputs, like in Duranton and Puga (2001)'s theory of nursery cities.

Spatial Decay Table (4) shows that higher block-level concentration of economic activity can lead to large productivity gains. Consistent with the idea that proximity matters, Table (6) shows that hyper-proximity maximizes productivity gains. Table (6) reports estimates from an specification identical to column (1) of Table (4), except that not only the characteristics of the incumbents in the same block are included but also the characteristics of the incumbents in all blocks within 300 meters of the block where the startup is located. These additional incumbents are included by rings. Figure (1) illustrate how these rings are constructed. The first ring includes all incumbents between 75 meters and 150 meters, the second ring all incumbents between 150 and 225 meters and the third ring all incumbents between 225 and 300 meters. The average employment and revenue in these cases is discounted by the inverse of the distance between the startup block and the incumbent blocks.

Estimates show that local spillovers primarily operate within 75 meters, and their effect quickly fades as moving away from the source of exposure. Estimates of the elasticities within the same block barely change relative to the baseline in Table (4). However, the elasticity quickly decays towards zero from 75 meters away in the case of average revenue and from

150 meters away in the case of employment. These results show that hyper-proximity to economic activity is essential for spillovers, proving that spatial decay results from Baum-Snow, Gendron-Carrier, and Pavan (2021) and Liu, Rosenthal, and Strange (2017) are generalizable to the rest of the economy. Moreover, these results provide an additional check for identification within neighborhoods. Including the surrounding blocks as controls account for sorting occurring across rings within the neighborhood. Obtaining persistent estimates for the same block exposure and non-significant estimates for the rings around it provides evidence that the economic neighborhoods are well constructed and that the spillovers are well identified.

[Table 6 about here.]

Heterogeneity Different industries and linkages across industries are likely to generate different levels and types of productivity spillovers. There is a long list of theories about the micro-foundations of agglomeration and how they relate to entrepreneurship, new firms, and their success (Chinitz, 1961; Jacobs, 1969; Marshall, 1890; Saxenian, 1994). These theories are about differentials in returns, the availability of some sort of input, and/or the environment. Separately identifying mechanisms has been a challenge that the rest of the literature and myself have faced, and I do not intend to overcome. Instead, I use measures that can be associated with one or another theory depending on the interpretation.

First, in the spirit of Faggio, Silva, and Strange (2017), I allow the elasticity to employment and revenue to differ across groups of industries while controlling for differences in market conditions that may arise across them. Table (7) reports estimates from an specification identical to column (1) of Table (4), except that the characteristics of the incumbents in the same block are interacted with indicator variables representing 5 groups of firms as in Glaeser, Kim, and Luca (2019). The interpretation of these coefficients are of elasticities across groups of industries. Given that exposure is given at a very small scale, it is very unlikely that there are differences in the environment across industries. In that case, what my estimates are capturing is a mix of differentials in returns and the availability of some sort of input (goods, services, knowledge). Results show that most of the industries have positive overall same-industry spillovers. However, depending on the size of the incumbents in a given location, the information and financial services and the manufacturing industries might not. Employment-intensive industries (such as retail, trade and manufacturing) benefit more from larger incumbent firms, while knowledge-intensive industries (such as professional, information and financial services) benefit more from higher quality incumbents.

[Table 7 about here.]

Second, as in Ellison, Glaeser, and Kerr (2010) I use cross-industry connectivity weights to study the importance of mechanisms underlying different theories of agglomeration. Table (8) reports estimates from an specification identical to column (1) of Table (4), except that incumbent characteristics are computed using cross-industry weights. This means that incumbent from all industries and not only incumbents from the same industry are included in this estimation. Column (1) reports results using equal weight to represent exposure to any kind of economic activity,[28] column (2) and (3) reports results using input-output weights to represent exposure to incumbents upstream and downstream the production process, and column (4) reports results using occupational similarity weights to represent knowledge transfer through workers performing similar occupations instead of firms performing similar activities.[29]

[Table 8 about here.]

Results using connectivity weights provide further evidence about the way productivity spillovers work for new firms. Regardless of differences in magnitudes across cross-industry weights, all estimates show positive spillovers between incumbent firms and startups. Qualitatively, these estimates support multiple agglomeration theories that recognize the importance of inputs for entrepreneurship (Chinitz, 1961; Jacobs, 1969; Marshall, 1890; Saxenian, 1994). Quantitatively, the use of cross-industry weights provides information about the relative importance of these theories at fine scales. For example, for input-output linkages, the elasticities to average employment are significantly smaller relative to other linkages and also relative to the elasticity to average revenue, which is consistent ideas that smaller independent incumbents encourage new firms as it would make it easier for new firms to find suppliers (Chinitz, 1961; Saxenian, 1994). However, in terms of magnitudes, the elasticity

[28] The baseline estimates essentially assign a weight equal to one to all incumbent firms in the same block that belong to the same industry as the startup, and zero otherwise. In this case, column (1) reports results assigning all incumbent firms in the same block a weight equal one.

[29] FollowingEllison, Glaeser, and Kerr (2010) and Baum-Snow, Gendron-Carrier, and Pavan (2021), I construct occupational similarity measures based on the 2002 US National Industry Occupation Employment Matrix that for each industry, gives the share of employees in each four digit occupation. The weights are constructed calculating the correlation between industries of the shares of employees in each four digit occupation. I then scale each correlation down such that it sums one for each industry.

to employment from input-output linkages is significantly smaller than exposure to same-industry activity, which again smaller than exposure from workers performing similar tasks. This is in line with knowledge flows that are more likely to occur between different industries with workers performing similar tasks than within the same industry, an idea that is consistent with spillovers depending not just on aggregate sizes but also on the nature of economic interactions that occur with proximity (Jacobs, 1969).[30]

4.2 Local Spillovers

The evidence I provide in this section delivers important insights about the nature and scale of productivity spillovers for new firms. On average, all types of new firms in my sample have important gains from concentration of economic activity that leads to better short and medium term economic outcomes. However, proximity to such concentration is very important as gains fade rapidly with distance. Moreover, there is stark heterogeneity in these results across industries and cross-industry linkages. These results have direct implications over the understanding of the role of economic conditions for new firms success, and showcase that there is no one policy that fits all.

These results use within neighborhood variation to identify local effects. This, the results apply directly only to differences in exposure occurring within neighborhoods. Moreover, the implications do not account for the important process underlying the location choice made by entrepreneurs. Recall that within neighborhoods, new firms are subject to vacancies and hence entrepreneurs do not have influence over which block within neighborhoods they will end up operating. The use of neighborhood fixed effects controls for fixed unobserved differences within the boundaries of the newly defined geographies. This allows me to control for sorting into blocks within neighborhoods. However, entrepreneurs do make a decision at the neighborhood level. Analyzing how the elasticities change while accounting for this decision informs about the importance of entrepreneur-driven sorting relative to location attributes for new firms. I attempt to do this with a entrepreneur location choice model that I develop in the next section.

[30] This type of weights have been used before to capture labor pooling. I interpret a worker moving from one firm to another and transferring knowledge through that movement as knowledge spillovers although it is technically part of the definition of labor pooling. I make this distinction because in such small geographies, labor pooling and knowledge spillovers can be operating through similar channels.

5 Local Spillovers and Neighborhood Sorting

To account for entrepreneurs sorting into neighborhoods, this section develops a two step model of neighborhood choice. The model builds on G. Dahl (2002), which proposes a semi-parametric estimation method for polychotomous choice models. The original model focuses on migration choices where self-selection arises from differentials in returns to education across locations within a country. In my framework, choices are limited to neighborhoods within a city and self-selection occurs as a consequence of entrepreneur-specific preferences for location. After an entrepreneur has chosen a neighborhood, real estate vacancies are realized and the startup is assigned to a block within that neighborhood. As in the previous section, these vacancies arise randomly within neighborhoods. This means that entrepreneurs do not perfectly observe outcomes at the time of the decision, and can only make decisions based on the expectation of outcomes at the neighborhood level.

As a preview, I estimate this model in two steps. The first step consists of estimating the probabilities of selection into neighborhoods. I identify these probabilities using personal preferences for location that I capture by allowing the probabilities to across entrepreneurs based on their demographics and distance to their previous place of residence. The second step uses these probabilities to construct correction functions that are included in the outcome equation as regressors. These correction functions account for self-selection into neighborhoods.

5.1 Model

In this section, I present a Roy model for multiple location choice that builds on G. Dahl (2002), adapting the analysis to startup neighborhood choices and introducing uncertainty on future outcomes through random block assignment within neighborhoods. For simplicity, I suppress time, industry, and city subscript the first subsections where I define the model. I add them back in the estimation subsection.

Consider I entrepreneurs facing N neighborhoods. In this stylized world there are two periods. In the first period, conditional on deciding to start to operate, entrepreneurs select a neighborhood following an expected utility maximization process that includes both pecuniary and non-pecuniary components. In the second period, once a block has been assigned, the firm starts operation and outcomes are realized. The model assumes that even though entrepreneurs have uncertainty about which block within a neighborhood they will be assigned, there is no uncertainty about expected outcomes or personal preferences across all neighborhoods. Hence, in the first period, each entrepreneur i compares the expected benefits obtainable in each of the neighborhoods and opts for the one that delivers the best bundle of expected-pecuniary and non-pecuniary benefits.

Formally, an entrepreneur's expected utility for neighborhood n is a function of two components: a pecuniary component given by their startup expected outcomes in that neighborhood, and a non-pecuniary component that represents the entrepreneur's neighborhood-specific location preferences:

$$V_n^i = \overline{y}_n^i + t_n^i. \tag{3}$$

In (3), V_n^i indexes utility, $\overline{y}_n^i$ is the expected outcome of the firm, and t_n^i is a vector indexing personal preferences for opening a firm in neighborhood n. This vector of personal preferences, t_n^i, encompasses all non-pecuniary utility components that could determine the utility of entrepreneur i starting operations in neighborhood n. These include local characteristics such as cultural and geographical amenities, personal connection with the neighborhood and commuting time.

In order to obtain expected outcomes, I use equation (1) to define startup potential outcome in block b of neighborhood n as,

$$y_{bn}^i = x^i\beta + X_{bn}\theta + \nu_{bn}^i, \tag{4}$$

in which y_{bn}^i denotes a firm outcome like revenues or employment, x^i are firm characteristics, X_{bn} denotes an aggregation of characteristics of incumbent firms, and ν_{bn}^i is an idiosyncratic term. These outcomes are not perfectly observed by the entrepreneur at the time of her decision, as they have uncertainty about the specific block within a neighborhood that they will be assigned. However, they know that assignment within neighborhoods is uniformly distributed. This means that entrepreneurs make location choices not considering the potential outcomes in a given block, but the expected potential outcomes in the neighborhood that contains that block. Taking expectations on equation (4) leads to,

$$\overline{y}_n^i \equiv \mathrm{E}_n^i \left[y_{bn}^i \mid x^i \right] \quad = \quad x^i\beta + \overline{X}_n^i\theta + \mathrm{E}_n^i \left[\nu_{bn}^i \mid x^i \right]$$

where $\overline{X}_n^i = \mathrm{E}_n^i\left[X_{bn}^i\right]$ corresponds to the unconditional mean of incumbents characteristics at the neighborhood level. The assumption of uniform assignment within neighborhoods has implications about the structure of the error term,

$$\mathrm{E}_n^i\left[\nu_{bn}^i \mid x^i\right] = \varepsilon_n^i$$

which shows that idiosyncratic shocks are distributed such that the expectation at the neighborhood level is idiosyncratic as well. Using this equation, the deviation of the expected outcomes of a startup if they were to start operating in neighborhood n from the average of the entire population is given by,

$$\overline{y}_n^i - \mathrm{E}\left[y_{bn}^i \mid x^i\right] = \varepsilon_n^i \tag{5}$$

This equation simply states that the deviation of entrepreneur expected outcomes from the average of other startups given her observable characteristics is the neighborhood specific component of the error term ν_{bn}^i in equation (4).

Without making specific assumptions about the personal preferences of entrepreneurs, I define similar equation for the deviation of entrepreneurs taste for location relative to the population average,

$$t_n^i - \mathrm{E}\left[t_n^i \mid z^i, d_n^i\right] = \mu_n^i \tag{6}$$

this implies that the expression for V_n^i can now be written in terms of population averages (or the so called sub-utility functions in the selection literature) and an error component specific to the entrepreneur-firm in a given location,

$$V_n^i = V_n + e_n^i \tag{7}$$

where $V_n^i = \mathrm{E}\left[y_{bn}^i \mid x^i\right] + \mathrm{E}\left[t_n^i \mid z^i, d_n^i\right]$ and $e_n^i = \varepsilon_n^i + \mu_n^i$. We can use this expression to define the following neighborhood selection process for firm-entrepreneur i,

$$\begin{aligned} M_n^i &= 1 \quad \text{if and only if} \quad V_n^i + e_n^i > V_k^i + e_k^i \quad \forall k \neq n \\ &= 0 \quad \text{otherwise} \end{aligned} \tag{8}$$

which means that outcomes of i are only observed for neighborhood n such that M_n^i=1, which is the neighborhood that maximizes the expected utility of the entrepreneur. Specifically, outcomes are only observed if all selection equations in (8) are simultaneously satisfied. Thus, firms observed operating in neighborhood n are not a random sample of the population; hence,

$$\begin{aligned} \mathrm{E}\left[\nu_{bn}^{i} | M_{n}^{i}=1\right] &= \mathrm{E}\left[\varepsilon_{n}^{i} | M_{n}^{i}=1\right] \\ &= \mathrm{E}\left[\varepsilon_{n}^{i} \mid e_{n}^{i}-e_{k}^{i} \leq V_{n}^{i}-V_{k}^{i} \quad \forall k \neq n\right] \\ &\neq 0 \end{aligned} \tag{9}$$

where $\mathrm{E}\left[\nu_{bn}^{i} | M_{n}^{i}=1\right]$ is the selection bias for i.

5.2 *Correcting for Neighborhood Sorting*

The most common procedure for estimation of models with self-selection and binary outcomes is the Heckman selection model. In polychotomous choice models, the selection correction term can be can be estimated via a conditional logit model or its extension the nested logit model which, after some transformations, can be included in the outcome equation to control for selection into choices (Vella, 1998).

All these methods rely on additional assumptions on the joint distribution of the error terms in the outcome and selection equations, as departures from the true joint distribution can lead to severe bias in the estimates as the size of the self-selected sample increases. G. Dahl (2002) developed a semi-parametric estimation approach motivated by the observation that in single-index selection models, the selectivity bias can be written as a function of the probability of selection given covariates (Ahn & Powell, 1993; Lee, 1983).

The intuition of G. Dahl (2002) method is as follows: the probability of observing a startup's outcome in block b of neighborhood n is related to the probability that V_n^i is the maximum of all sub-utility functions. Thus, the joint distribution between the error term in the earnings equation (ε_{bn}^i) and the differenced sub-utility error terms $\{e_n^i - e_k^i\}_{\forall k \neq n}$ can be reduced from N dimensions to two dimensions: the first dimension is the outcomes error and the second is the maximum order statistic of the differenced sub-utility functions. The key assumption is that this bivariate distribution does not depend on the sub-utility functions themselves, except through a small number of choice probabilities. This allows the researcher to express the selection correction term in the earnings equation (analogous to the inverse Mills ratio term in the canonical Heckman selection model) as a function of a small number of observed choice probabilities. Without this assumption, the researcher would be required to estimate an $(N-1)$-dimensional integral. This becomes quickly infeasible as N grows large, as is the case in the current setting.

Formally, the selection rule in (8) implies that,

$$y_n^i \qquad \text{is observed iff} \quad \max_{k \neq n} \left(V_n^i - V_k^i + e_n^i - e_k^i \right) \leq 0 \tag{10}$$

where $\max_k(\bullet)$ correspond to the maximum over all neighborhoods $k \neq n$. Thus, any selection bias in y_n^i is driven by the probability event that the maximum of the collection of random variables $\{V_n^i - V_k^i + e_n^i - e_k^i\}_{k \neq n}$ is less or equal to zero. This means that when the distribution of this maximum is known, selection bias can be controlled in the same way as in a binary choice models under some assumptions.

In particular, if we believe that the distribution of the error term in the outcome equation, $\varepsilon_n^{(i)}$, is normal and that the distribution of the random portions of the indirect utility function, $\{e_n^i - e_k^i\}_{\forall k \neq n}$, are assumed to be independent and identically distributed with the extreme value distribution. Then, a transformation of the maximum order statistic to normality would allow us to assume a joint bivariate normal distribution for the error term in the outcome equation and the transformed maximum order statistic. Then, a simple Heckman-type correction will control for selectivity bias, by adding a term that takes the form of the inverse Mill's ratio to the outcome regression function.

The formulation of neighborhood selection and firm outcomes in equations (4) and (10) implies that outcomes equations can be rewritten as multiple-index, partially linear models:

$$y_{bn}^i \quad = \quad x^i\beta \ + \ X_{bn}\theta \ + \ M_n^i \ \times \ \Omega_n\left(V_1^i - V_n^i, V_2^i - V_n^i \ldots, V_N^i - V_n^i\right) \ + \ \mu_{bn}^i \tag{11}$$

where $\Omega_n(\bullet) = \mathrm{E}_n\left[e_n^i | V_1 - V_n, V_2 - V_n \ldots, V_N - V_n\right]$, μ_n^i is an error term with mean zero in the conditional sample for neighborhood n, and M_n^i is a dummy variable that equals one if firm i chooses neighborhood n. By controlling for all sub-utility differences through unknown equations Ω_k^N for every possible neighborhood k, this multiple index equation successfully controls for neighborhood self-selection. The caveat of this equation though is that it brings back the problem of high-dimensionality. However, we can combine this equation with Lee (1983)'s maximum order statistic approach to reduce it by making the following index sufficiency assumption,

$$g_n\left(\varepsilon^i_n, \max_{k\in N}\left(V^i_n - V^i_k + e^i_n - e^i_k\right)|V_1 - V_n, V_2 - V_n \ldots, V_N - V_n\right)$$
$$= g_n\left(\varepsilon^i_n, \max_{k\in N}\left(V^i_n - V^i_k + e^i_n - e^i_k\right)|\mathrm{P}^i_n\right) \quad (12)$$

where g_n is the density function of a well defined joint distribution function for the error term in the outcome equation and the maximum order statistic and P^i_n is the probability that firm i chooses neighborhood n given $\{V^i_n - V^i_k + e^i_n - e^i_k\}_{\forall k\neq n}$. This assumption means that all the information about all potential alternative neighborhood is included in the probability that an entrepreneur chooses one specific neighborhood. Under this assumption, equation (11) can be simplified to a single index equation,

$$y^i_{bn} = x^i\beta + X_{bn}\theta + M^i_n \times \lambda_n\left(\mathrm{P}^i_n\right) + \omega^i_{bn} \quad (13)$$

where for each neighborhood n, $\lambda_n(\bullet)$ is an unknown function of the single index P^i_n, and ω^i_{bn} is an error term. Providing there is a good estimator for P^i_n that accounts for all sub-utility functions, $\lambda(\bullet)$ corrects for self-selection into neighborhood n since by construction the error term $\omega^{(i)}_{nt}$ has zero mean given the selection probability, the fact that the firm is observed in that neighborhood and observables.

Even though the index sufficiency assumption allow us to decrease the dimensionality of the problem and correct for self selection into neighborhoods, it is not without strong restrictions. Index sufficiency holds, for example, if outcome errors are composed of a firm level fixed effect that is invariant to entrepreneur characteristics including their place of residence. On the other hand, this assumption is less likely to hold in a setting where, for example, there is a correlation between that fixed effect with location characteristics. This is important to reconcile the fact that as showed in (2.2), entrepreneurs tend to open their firms at their place of residence disproportionally more than in other neighborhoods across the city. I incorporate this into the construction of the correction functions by allowing that the correction function varies across entrepreneur characteristics and not only including the first best probability but also the probability to start operations in their place of residence. This acts as an exclusion restriction that discuss in the next section together with some technical aspect of the estimation.

5.3 Estimation details

In this section, I discuss details about the estimation and identification of equation (13). I first discuss how I separately identify firm outcomes from personal preferences for location. Then, I detail the estimation process that, as in other implementations of similar models, consist of two stages. First, I estimate neighborhood choice probabilities. Second, I use these probabilities to construct correction functions that I correct the error term of (4) and allows me to obtained unbiased estimates by estimating equation (13) using ordinary least squares.

Identification

As discussed in other implementations of the extended Roy model (Bayer, Khan, & Timmins, 2011; G. Dahl, 2002; D'Haultfœuille & Maurel, 2013; Ransom, 2021), separately identifying outcomes from non-pecuniary preferences requires an exclusion restriction—a covariate which appears in the neighborhood choice probabilities but does not affect startup outcomes.

Section (2.2) showed that not only do entrepreneurs tend to open their firms at their place of residence disproportionally more than in other neighborhoods across the city, but that in the case of deciding to open in another location, they also do so closer to their place of residence. Moreover, it shows that entrepreneurs of different demographics make different neighborhood choices regardless of location characteristics.

This fact is heavily related to the premise of the paper and is the root of the exclusion restriction I use to separately identify firm outcomes from personal preferences for location. Where to open a new firm is one of the most important decisions entrepreneurs have to face, and that often this decision is not only influenced by the location potential for the new firm but also by personal preferences. Many entrepreneurs are willing to substitute some firm profit for benefits of living in attractive places or places to which they have some sort of personal connection (M. Dahl & Sorenson, 2012; Figueiredo, Guimarães, & Woodward, 2002; Kulchina, 2016). Preferences also influence investment decisions, as home bias in investment portfolios (Goetzmann & Kumar, 2008).

The primary threat to the validity of the exclusion restriction is if the neighborhood decision of the entrepreneur is driven by advantageous draws of the outcome equation at the place of residence. In other words, there are unobservable productivity benefits of opening in the place of residence of one of the owners. Even if there might be some benefits of starting a new firm in the place of residence, these are likely to be in the form of cheaper rents which are assumed to be constant across the city and affect all entrepreneurs in the same industry in the same way. Moreover, regardless of that fact, there is a strong correlation across industries and cities in the propensity to start operations.

The primary threat to the validity of the exclusion restriction is that firms that choose to start operations away from the place of residence are doing so by firm specific benefits of agglomeration economies. This is related to Baum-Snow, Gendron-Carrier, and Pavan (2021) and Guzman (2019) that finds that "movers" are more likely to be high quality firms. Even though this might be possible, evidence in Section (2.2) shows owners of different demographics (likely unrelated to productivity, such as gender, age and immigrant status) make different location choices. Hence, the propensity to open in the place of residence, or any place for that matter, appears to be in part driven by entrepreneur personal preferences.

Choice Probabilities

The formulation of equation (13) assumes the researcher possesses consistent estimates of the relevant neighborhood choice probabilities. In response to the known drawbacks of models that require strong distributional assumptions (like the conditional logit model), researchers like G. Dahl (2002) and Ransom (2021), have used non parametric estimators of the probabilities. This estimators have the advantage that it allows the researcher to not model the way the sub-utilities depend on location-specific characteristics. Moreover, it allows them to avoid making strong distributional assumptions.

However, data requirement is an important limitation that forces researchers to compromise representativity of the estimates in order to obtain cells of data that contains enough variation to compute the probabilities.

The objective of this paper does not rely on the separate identification of personal preferences from firms outcomes. It only focuses on cleaning up the selection bias produced by the effect that personal preferences have on location choices. This distinction allows me to treat this as a classification problem, in which given entrepreneurs/location characteristics it assigns a probability to each location providing they maximize the fit of the estimated probabilities to the observed location choices.

Classification problems are common in computer science applications in which methods are created to provide quick and accurate results with large amounts of data, and more importantly for this case, very large choice sets. In particular, neural networks models allows for data-oriented modifications from the text-book McFadden (1973) model to be able to handle not only large choice sets but also large amounts of independent variables (or choice features as called in the machine learning literature). This implies that I can relax the constraint of making parametric decisions about the set of variables that affect the sub-utilities by adding a regularization parameter that penalizes the number of variables included.

In particular, neural networks are mathematical tools that are loosely inspired by the functional aspects of biological neural systems. Biological neural systems consist of multiple nodes, called neurons, that communicate through synapses. Typically, there are three sets of nodes: Input nodes, intermediate (hidden) nodes, and output nodes, and each node category plays a different role. Input nodes receive input information, output nodes yield output signals, and intermediate nodes receive signals from input nodes, and manipulate those signals to give results to output nodes. Conditional choice models share similar features, in which choice-specific characteristics (input note) are translated (intermediate node) into choice-specific probabilities (output node).

This characterization is useful, as it means I can use neural networks to compute entrepreneur-specific probabilities at the neighborhood level. This is done treating each individual as a unique node in the network. Each node provides information about the relationship between features and choices and informs the rest of the network through forward and backward loops. This is done in sequence, in which each node multiplies input values by a weighting vector, adds a bias, and then applies a non-linear activation function like a logit function (also called softmax). Outputs of each node are propagated (and back-propagated) through the network which realizes the true power of neural networks - the ability of early layers to learn representations that can be utilized by later layers in the network.

This process is done using algorithms that optimize the value of an objective function such as a likelihood function. In a seminal paper, Kingma and Ba (2017) proposed the Adam algorithm, which has quickly become the standard in computer science as it achieves unbiased estimates fast. Adam is an adaptive learning rate method, which means, it computes individual learning rates for different parameters. It does so using the estimations of first and second moments of gradient to adapt the learning rate for each weight of the neural network. It then uses the learning rates to update the weights (or parameters) that govern the relationship between choice characteristics and probabilities, and then activates the node by calculating the loss rate. This is done a maximum amount of times (iterations or epochs in the machine learning literature) or until convergence, whichever comes first. Standard errors of estimated parameters are equal to the square root of the inverse of Hessian of the log-likelihood at the optimal point.

I use this algorithm to compute choice-specific probabilities. My implementation relies on the fact that in order to apply the algorithm off the shelve and being able to calculate individual-choice specific probabilities we need to provide the algorithm with data at the chooser-choice level. Entrepreneurs in my setting face hundreds if not thousand of possible neighborhoods within a city which means creating such dataset ex-ante is not feasible.[31]

Table (9) shows descriptive statistics of the estimated probabilities using entrepreneur demographics and distance to their first residential neighborhood as an instrument for neighborhood choices. In my data, chosen neighborhoods have a mean rank of 31.29 out 189.6 available neighborhoods to choose from. This is consistent with estimates found in Qian and Tan (2021) that showcase a similar trend and imply that there might be important unobservables that I cannot account for with my specification. If these unobservables are correlated with sorting into neighborhoods, this is a threat to identification. However, the flexibility of the neural network approach reduces the importance of this problem relative to other approaches, such as a nonparametric estimation, as estimated probabilities are entrepreneur-specific.[32]

[Table 9 about here.]

Correction functions

After computing entrepreneur-neighborhood level probabilities, I use them to compute correction functions for each entrepreneur and neighborhood. Correction functions enter the outcome variable as regressors and allows unbiased ordinary least square estimates.

[31] I develop an implementation of the algorithm based in the python package `pytorch`. I modify the off-the-shelf algorithm to do this in the backend and developed a Stata program that allowed me to integrate it with the whole workflow of the rest of the paper.I call this package `nmlogit` or neural multinomial logit and it is in continuing development. It can be found in https://github.com/rcampusa/nmlogit. Suggestions and comments are heavily appreciated.

[32] Table (9) highlights interesting sorting patterns. Besides the fact that chosen neighborhoods are often below in the ranking of choices, the average of estimated probability for opening a firm in the place of residence is matched with the empirical probability of opening at home. Moreover, it shows that even though the probability is high, the ranking is lower than the chosen neighborhood but with higher variance.

Using neural networks for computing probabilities allows for probabilities that are entrepreneur-location specific. In previous work, researches have needed to make assumptions about how different (or equal in their case) the correction functions are across individuals. This assumption allowed them to compute estimates of the probabilities for group of people instead of individuals themselves. I don't need to make such assumption.

Given that entrepreneurs tend to open their firms at their places of residence disproportionally, the index sufficiency assumption in equation (12) likely invalid. This is because the probability of the first best does not necessarily include all the information used in the selection process. To account for this, I also control for the probability associated with their neighborhood of residence, which should account for the residual information left by the first best.

5.4 Local Spillovers and Location Sorting

The estimates in Section 4 are obtained using the within-neighborhood variation. Moreover, they are based on startups that decide to open away from entrepreneur places of residence. Table (2) in section (2) showed that an important portion of new firms decide to start operating in their places of residence. Moreover, it shows that different owners make different location choices, which implies that sorting is important in this setting. I now discuss the parameter estimates of the outcomes equation with and without the selection correction and using the full sample of startups.

In principle, standard errors of the parameters associated with the selection functions should be adjusted to account for the fact that they are estimated which might bias the coefficients of the correction functions as the true probabilities are not observed. Unfortunately, the implications for inference of using a model selection algorithm in the first stage are not well understood, so it is not clear how to proceed (Mullainathan & Spiess, 2017).

To estimate the unknown correction functions, I depart from using a flexible polynomial function of the probabilities as discussed in G. Dahl (2002). Previous studies show that including a higher degree polynomial or a larger number of probabilities results in much less precise estimates with no appreciable increase in the Wald test statistic of joint significance of the polynomial. Instead, as $\delta_{nt}^{(j)}$ is very high dimensional, interacting it with the estimated probabilities is flexible enough to account for sorting patterns occurring across neighborhoods. The updated estimating equation for startup i's operating in block b of neighborhood n at time t is

$$y_{bnt}^{(i,j)} = x^i\beta + X_{bnt}^{(j)}\theta + \delta_{nt}^{(j)} \times \left(\overset{star^{(i,j)}}{\mathrm{P}}_{nt}, \overset{home^{(i,j)}}{\mathrm{P}}_{nt} \right) + \omega_{bn}^{(i,j)} \tag{14}$$

where the fixed effect $\delta_{nt}^{(j)}$ is now interacted with either the first best probability, $\overset{star^{(i,j)}}{\mathrm{P}}_{nt}$, or the probability of opening in the place of residence, $\overset{home^{(i,j)}}{\mathrm{P}}_{nt}$.

Table (10) reports the results for the estimates of block-level spillovers while accounting for sorting within and across neighborhoods. Estimates show that accounting for sorting into neighborhoods increases the elasticities to average incumbent employment between 25 to 50%, while it decreases the elasticities to average revenue between 30 to 40%. Depending on the specification, differences between these estimates are slightly significant, which indicates that there even though there is sorting across neighborhoods its effects over the estimates is inconclusive. However, the signs of the differences provides insights that are consistent with the evidence within neighborhoods and that was highlighted throughout the rest of the paper. Startups benefit more from exposure to larger firms than from from exposure to smaller firms with higher revenue. This indicates that, on average, scale effects are more important for new firms early stages.

[Table 10 about here.]

6 Conclusions

This paper is about entrepreneurs, their startups, and the spatial scale of productivity spillovers. A substantial amount of evidence for regions and cities shows that firm and worker productivity increases in proximity to a high concentration of economic activity. This paper provides evidence that productivity spillovers apply to very local environments within cities and positively affect new firms' outcomes in the short and medium-term. However, proximity to such concentration is essential as gains fade rapidly with distance. Using a structural model, I find scant evidence of sorting across neighborhoods within cities. Accounting for sorting accentuates the patterns observed within neighborhoods. On average, firms benefit more from exposure to larger firms than from exposure to firms with higher sales. Given the importance of new firms in the economy, this type of study informs public policy and private efforts that aim to concentrate economic activity. This paper provides evidence that suggests that any type of effort that encourages concentration of economic activity should do it at very local levels and do it differently across industries.

Results are remarkably heterogeneous across industries and types of exposure. I use a newly developed algorithm to account for differences in the sorting patterns within neighborhoods across industries. I do this by recognizing that what one industry may find desirable, another industry might not. This allows pooling the data across industries with less concerns about the external validity of the results. Results are consistent with productivity spillovers operating differently across industries and showcase that even though the average firm benefits more from exposure to larger firms, this is not the case for all industries. Employment-intensive industries benefit relatively more from larger incumbent firms, while for knowledge-intensive industries is the opposite. Moreover, results from cross-industry linkages show that these spillovers are consistent with several theories of agglomeration.

References

Agarwal, R., & Moeen, M. (2016). Entrepreneurial Startups (de novo), Diversifying Entrants (de alio) and Incumbent Firms. In M. Augier & D. J. Teece (Eds.), *The Palgrave Encyclopedia of Strategic Management* (pp. 1–4). Palgrave Macmillan UK. (Cit. on p. 9).

Ahn, H., & Powell, J. L. (1993). Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics*, *58*(1-2), 3–29 (cit. on p. 29).

Alcácer, J., & Chung, W. (2014). Location strategies for agglomeration economies. *Strategic Management Journal*, *35*(12), 1749–1761 (cit. on p. 1).

Arzaghi, M., & Henderson, J. V. (2008). Networking off Madison Avenue. *The Review of Economic Studies*, *75*(4), 1011–1038 (cit. on pp. 1, 3, 12, 13).

Athey, S., & Imbens, G. (2019, March 24). *Machine Learning Methods Economists Should Know About*. arXiv: 1903.10075 [econ, stat]. (Cit. on p. 6).

Bailey, M., Gupta, A., Hillenbrand, S., Kuchler, T., Richmond, R., & Stroebel, J. (2021). International trade and social connectedness. *Journal of International Economics*, *129*, 103418 (cit. on p. 20).

Baldwin, J., Dupuy, R., & Penner, W. J. (1992). Development of Longitudinal Panel Data From Business Registers. *Statistical Journal of the United Nations*, *9*(4), 289–303 (cit. on p. 9).

Baldwin, J. R., Landry, L., & Leung, D. (2016). The Measurement of Firm Entry in the Longitudinal Employment Analysis Program. *Analytical Studies: Methods and References*, *11-633-X*(004), 35 (cit. on p. 8).

Baum-Snow, N., Gendron-Carrier, N., & Pavan, R. (2021). Local Productivity Spillovers. *mimeo*, 25 (cit. on pp. 23, 24, 33).

Bayer, P., Khan, S., & Timmins, C. (2011). Nonparametric Identification and Estimation in a Roy Model With Common Nonpecuniary Returns. *Journal of Business & Economic Statistics*, *29*(2), 201–215 (cit. on p. 32)
_eprint: https://doi.org/10.1198/jbes.2010.08083.

Campusano, R. (2021). *Delineating Neighborhoods using Location Choices*. (Cit. on pp. 4, 13, 15–17).

Chatterji, A. (2009). Spawned with a silver spoon? entrepreneurial performance and innovation in the medical device industry. *Strategic Management Journal*, *30*(2), 185–206 (cit. on p. 1).

Chatterji, A., Glaeser, E., & Kerr, W. (2014). Clusters of Entrepreneurship and Innovation. *Innovation Policy and the Economy* (Josh Lerner and Scott Stern). University of Chicago Press. (Cit. on p. 1).

Chinitz, B. (1961). Contrasts in agglomeration: New york and pittsburgh. *The American Economic Review*, *51*(2), 279–289 (cit. on pp. 23, 24).

Cochran, W. G. (1968). The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies. *Biometrics*, *24*(2), 295–313 (cit. on p. 16).

Combes, P. P., Duranton, G., Gobillon, L., Puga, D., & Roux, S. (2012). The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection. *Econometrica*, *80*(6), 2543–2594 (cit. on p. 1).

Correia, S., Guimarães, P., & Zylkin, T. (2020). Fast Poisson estimation with high-dimensional fixed effects. *The Stata Journal*, *20*(1), 95–115 (cit. on p. 20).

Dahl, G. (2002). Mobility and the Return to Education: Testing a Roy Model with Multiple Markets. *Econometrica*, *70*(6), 2367–2420 (cit. on pp. 5, 26, 29, 32, 33, 36).

Dahl, M., & Sorenson, O. (2012). Home Sweet Home: Entrepreneurs' Location Choices and the Performance of Their Ventures. *Management Science*, *58*(6), 1059–1071 (cit. on pp. 4, 32).

de Bellefon, M.-P., Combes, P.-P., Duranton, G., Gobillon, L., & Gorin, C. (2019). Delineating urban areas using building density. *Journal of Urban Economics*, 103226 (cit. on p. 17).

Decker, R., Haltiwanger, J., Jarmin, R., & Miranda, J. (2014). The Role of Entrepreneurship in US Job Creation and Economic Dynamism. *Journal of Economic Perspectives*, *28*(3), 3–24 (cit. on p. 8).

D'Haultfœuille, X., & Maurel, A. (2013). Inference on an extended Roy model, with an application to schooling decisions in France. *Journal of Econometrics*, *174*(2), 95–106 (cit. on p. 32).

Dixon, J., & Rollin, A.-M. (2012). Firm Dynamics: Employment Growth Rates of Small Versus Large Firms in Canada. *The Canadian Economy in Transition Series*, *11-622-M*(025), 41 (cit. on p. 8).

Duranton, G., & Puga, D. (2001). Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products. *American Economic Review*, *91*(5), 1454–1477 (cit. on p. 22).

Ellison, G., Glaeser, E. L., & Kerr, W. R. (2010). What Causes Industry Agglomeration? evidence from Coagglomeration Patterns. *The American Economic Review*, *100*(3), 1195–1213 (cit. on pp. 2, 24).

Faggio, G., Silva, O., & Strange, W. C. (2017). Heterogeneous Agglomeration. *Review of Economics and Statistics*, *99*(1), 80–94 (cit. on pp. 1, 2, 23).

Figueiredo, O., Guimarães, P., & Woodward, D. (2002). Home-field advantage: Location decisions of Portuguese entrepreneurs. *Journal of Urban Economics*, *52*(2), 341–361 (cit. on pp. 4, 32).

Franco, A. M., & Filson, D. (2006). Spin-outs: Knowledge diffusion through employee mobility. *The RAND Journal of Economics*, *37*(4), 841–860 (cit. on p. 9).

Fujita, M., Krugman, P., & Venables, A. J. (2001, July 27). *The Spatial Economy: Cities, Regions, and International Trade*. The MIT Press. (Cit. on p. 1).

Glaeser, E. L., Kerr, S. P., & Kerr, W. R. (2015). Entrepreneurship and Urban Growth: An Empirical Assessment with Historical Mines. *Review of Economics and Statistics*, *97*(2), 498–520 (cit. on p. 1).

Glaeser, E. L., Kim, H., & Luca, M. (2019). Nowcasting the Local Economy: Using Yelp Data to Measure Economic Activity. *NBER Chapters*. National Bureau of Economic Research, Inc. (Cit. on p. 23).

Glaeser, E. L., & Maré, D. C. (2001). Cities and Skills. *Journal of Labor Economics*, *19*(2), 316–342 (cit. on p. 22).

Goetzmann, W. N., & Kumar, A. (2008). Equity Portfolio Diversification. *Review of Finance*, *12*(3), 433–463 (cit. on p. 32).

Greenstone, M., Hornbeck, R., & Moretti, E. (2010). Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings. *Journal of Political Economy*, *118*(3), 536–598 (cit. on pp. 1, 3, 12, 13).

Guimarães, P., Figueirdo, O., & Woodward, D. (2003). A Tractable Approach to the Firm Location Decision Problem. *Review of Economics and Statistics*, *85*(1), 201–204 (cit. on p. 15).

Guzman, J. (2019, March 1). *Go West Young Firm: Agglomeration and Embeddedness in Startup Migrations to Silicon Valley* (SSRN Scholarly Paper No. ID 3175328). Social Science Research Network. Rochester, NY. (Cit. on p. 33).

Haltiwanger, J., Jarmin, R. S., & Miranda, J. (2012). Who Creates Jobs? small versus Large versus Young. *The Review of Economics and Statistics*, *95*(2), 347–361 (cit. on p. 8).

Heblich, S., Seror, M., Xu, H., & Zylberberg, S. Y. (2019, May 21). *Industrial clusters in the long run: Evidence from Million-Rouble plants in China* (No. 19/712). School of Economics, University of Bristol, UK. (Cit. on pp. 3, 4, 12, 13).

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, *47*(1), 153–161 (cit. on p. 5).

Henderson, J. V., Kuncoro, A., & Turner, M. (1995). Industrial Development in Cities. *Journal of Political Economy*, *103*(5), 1067–1090 (cit. on p. 1).

Jacobs, J. (1969). *The economy of cities*. New York: Vintage. (Cit. on pp. 23–25).

Kingma, D. P., & Ba, J. (2017, January 29). *Adam: A Method for Stochastic Optimization*. arXiv: 1412.6980 [cs]. (Cit. on pp. 5, 34).

Kulchina, E. (2016). Personal Preferences, Entrepreneurs' Location Choices, and Firm Performance. *Management Science*, *62*(6), 1814–1829 (cit. on pp. 4, 32).

Lee, L.-F. (1983). Generalized Econometric Models with Selectivity. *Econometrica*, *51*(2), 507–512 (cit. on pp. 5, 29, 30).

Liu, C. H., Rosenthal, S. S., & Strange, W. C. (2017). *Building Specialization, Anchor Tenants and Agglomeration Economies*. (Cit. on pp. 12, 13, 23).

Liu, C. H., Rosenthal, S. S., & Strange, W. C. (2018). The vertical city: Rent gradients, spatial structure, and agglomeration economies. *Journal of Urban Economics*, *106*, 101–122 (cit. on p. 3).

Marshall, A. (1890). *Principles of Economics*. Macmillan and Company. (Cit. on pp. 23, 24).

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, 105–142 (cit. on pp. 14, 34).

Mullainathan, S., & Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives*, *31*(2), 87–106 (cit. on pp. 6, 36).

Nathan, M., & Overman, H. (2013). Agglomeration, clusters, and industrial policy. *Oxford Review of Economic Policy*, *29*(2), 383–404 (cit. on p. 1).

Porter, M. E. (1998). *Clusters and the new economics of competition* (Vol. 76). Harvard Business Review Boston. (Cit. on p. 1).

Qian, F., & Tan, R. (2021). *The Effects of High-skilled Firm Entry on Incumbent Residents*. (Cit. on pp. 3, 4, 12, 13, 35).

Ransom. (2021). Selective Migration, Occupational Choice, and the Wage Returns to College Majors. *Annals of Economics and Statistics*, (142), 45 (cit. on pp. 32, 33).

Roche, M. P. (2020). Taking Innovation to the Streets: Microgeography, Physical Structure, and Innovation. *The Review of Economics and Statistics*, *102*(5), 912–928 (cit. on pp. 3, 12, 13).

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing Bias in Observational Studies Using Subclassification on the Propensity Score. *Journal of the American Statistical Association*, *79*(387), 516–524 (cit. on p. 15).

Rosenthal, S. S., & Strange, W. C. (2001). The Determinants of Agglomeration. *Journal of Urban Economics*, *50*(2), 191–229 (cit. on p. 1).

Rosenthal, S. S., & Strange, W. C. (2003). Geography, Industrial Organization, and Agglomeration. *Review of Economics and Statistics*, *85*(2), 377–393 (cit. on p. 1).

Rosenthal, S. S., & Strange, W. C. (2004). Evidence on the Nature and Sources of Agglomeration Economies. In J.-F. Thisse & J. V. Henderson (Eds.), *Handbook of Regional and Urban Economics* (pp. 2119–2171). Elsevier. (Cit. on p. 1).

Roy, A. D. (1951). Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers*, *3*(2), 135–146 (cit. on p. 4).

Saxenian, A. (1994). *Regional Advantage: Culture and Competition in Silicon Valley and Route 128, With a New Preface by the Author*. Harvard University Press. (Cit. on pp. 23, 24).

Schoar, A. (2010). The Divide between Subsistence and Transformational Entrepreneurship. *Innovation Policy and the Economy*, *10*, 57–81 (cit. on p. 8).

Silva, J. M. C. S., & Tenreyro, S. (2006). The Log of Gravity. *The Review of Economics and Statistics*, *88*(4), 641–658 (cit. on p. 20).

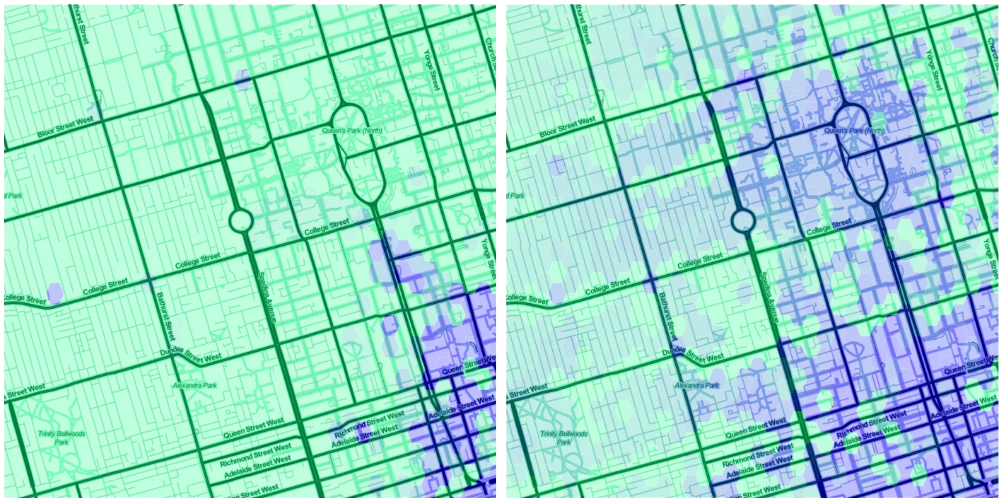
Vella, F. (1998). Estimating Models with Sample Selection Bias: A Survey. *The Journal of Human Resources*, *33*(1), 127 (cit. on p. 29).

Figures

Figure 1: Illustration of Hexagon Grid

Note: Each firm is assigned to an hexagon of 75 meter sides. Incumbents characteristics are computed at the hexagon level. The area covered by one hexagon is similar to a city block. Because of that, in the rest of the paper I refer to hexagon cells and blocks interchangeably. The rings highlighted around the center grid will be used to study spatial decay.

Figure 2: Spatial Distribution of Propensity Score



(a) Manufacturing (NAICS 3)

(b) Entertainment and Food Services (NAICS 7)

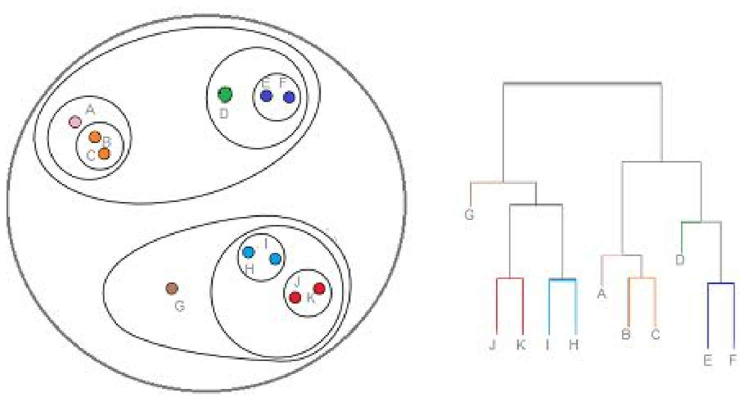
Notes: This figure shows the spatial distribution of the predicted probability for the manufacturing and entertainment industries in a one kilometer radius around a major intersection in the city of Toronto. To maintain confidentiality of CEED data, these probabilities are estimated using only the external DMTI data.

Figure 3: What is a Neighborhood?

(a) Postal Codes (6-digit Postal Codes) (b) Local Delivery Units (3-digit Postal Codes)

Notes: This figure superpose 3 and 6 digit postal codes to the previous figures to show that the spatial distribution of the predicted probability is not homogeneous within 3-digit postal codes and present high spatial correlation across 6-digit postal codes, which poses an identification concern about using postal code fixed effects to account for sorting.

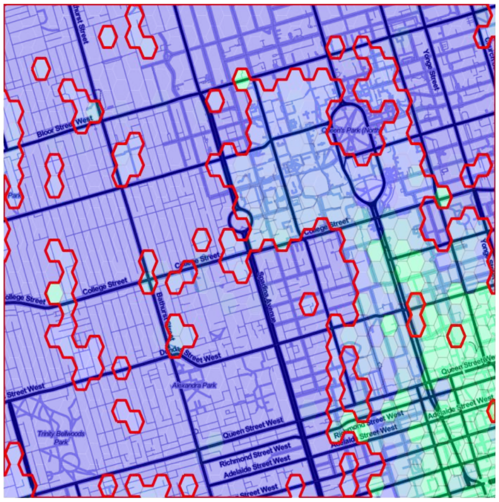
Figure 4: Illustration of the Algorithm



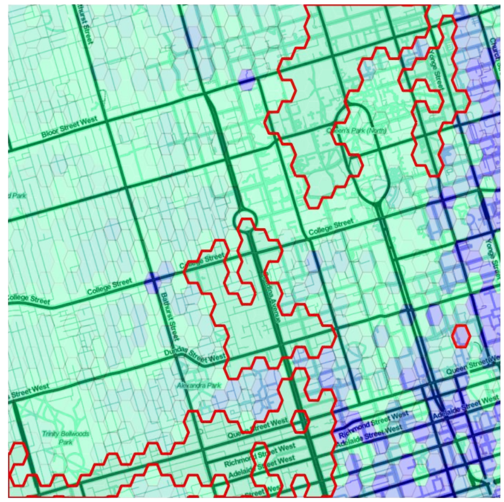
The algorithm starts with each block as a neighborhood by itself. Further iterations merge one neighborhood at a time so that the bilateral distance between the neighborhood and the new elements are below a given threshold and together minimize the within neighborhood variance in the propensity score.

Figure 5: Economic Neighborhoods

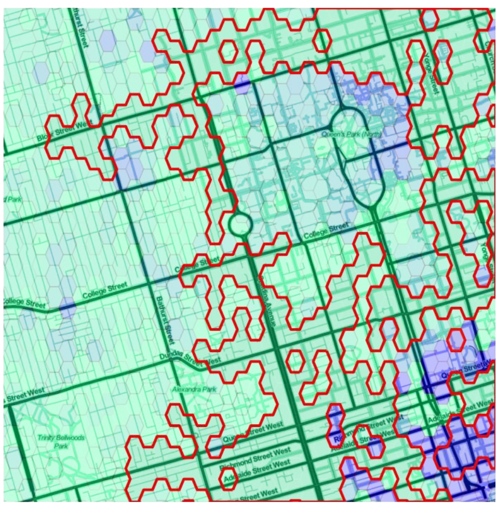
(a) Manufacturing



(b) Retail



(c) Professional Services



(d) Entertainment and Food Services

Notes: Each figure shows the example of resulting economic neighborhoods and how they are different around the same intersection across industries. To maintain confidentiality of CEED data, the estimated probabilities and neighborhoods in these maps are the result of using only the external DMTI data.

Tables

Table 1: Descriptive Statistics of Firms

	Incumbent	**Startup**	**Inst. Entrant**
Revenue (in millions, 2000 CAD)	2.245	0.158	1.704
Employment	6.73	0.75	3.39
Alive within 5 years	0.862	0.424	0.351
# Firms	1,347,920	317,606	248,364
# Firms Years	7,873,676	317,606	248,364

Note: Statistics are for all firms in the Calgary, Montreal, Toronto and Vancouver CMAs for the 2007-2018 period. Each statistic is the mean across their type of firms. Startups and institutional entrants are only accounted for during their founding year, incumbents are during all years.

Table 2: Descriptive Statistics of Firm Locations

	Incumbents	**Startups**	**Inst. Entrant**
Located in residence of a founder*	0.446	0.633	0.821
Move within 5 years	0.567	0.679	0.652
Firms in Same Industry	6.846	5.256	11.18
Average Revenue Employment Same Industry	14.739	6.516	22.821
Average Revenue Incumbents Same Industry	3,241,941	1,135,098	3,831,553
Number of Institutional Entrants	2.704	3.597	5.858
Number of New Startups	4.023	7.734	4.599

Notes: This table reports means of block-level characteristics for the location of the three types of firms. For startups and inst. entrants, these characteristics correspond to the founding year. For incumbents, these characteristics correspond to the average along all years. *In the case of Inst. Entrants, the percentage of firms located in residence of a founder is given relative to the place of operations of the parent company. In the case of incumbents is given by either the residence of individual owners or place of operations of the parent company.

Table 3: Number of Startups in a Block

	Toronto	Vancouver	Montreal	Calgary
Manufacturing	768.6	-233.6	982.8**	858.8
	(0.55)	(-0.32)	(2.86)	(1.31)
Wholesale Trade	-1789.4	-1683.4	411.1	1151.1
	(-1.02)	(-1.60)	(1.01)	(1.74)
Retail Trade	1358.1	-2529.3	991.0	4255.4***
	(0.63)	(-1.20)	(1.81)	(3.69)
Transportation	7878.5	-1075.6	972.3	951.3
	(1.93)	(-0.97)	(1.46)	(0.38)
Information	2503.0	719.2	655.0**	-316.6
	(1.71)	(0.77)	(2.62)	(-0.59)
Finance and Insurance	1686.3	998.6	1859.7**	-61.76
	(0.96)	(0.91)	(3.23)	(-0.06)
Real Estate	6893.2**	-184.3	2569.1***	-737.4
	(3.27)	(-0.11)	(4.01)	(-0.52)
Professional, scientific and technical services	5154.8	-2067.6	4080.9***	-1220.2
	(0.67)	(-0.77)	(3.31)	(-0.19)
Management of companies and enterprises	1867.9	564.6	661.9**	759.6
	(1.83)	(1.08)	(2.72)	(1.28)
Support, Waste Management and Remediation	3753.4*	247.1	1268.9*	1203.6
	(2.33)	(0.27)	(2.00)	(0.75)
Arts, Entertainment and Recreating	2580.0*	1083.6*	172.4	-216.7
	(2.37)	(2.32)	(0.83)	(-0.44)
Accommodation and Food services	2352.0	2783.3*	336.8	761.9
	(1.20)	(2.30)	(0.75)	(0.84)

Notes: This table shows the correlation between the contemporaneous propensity score with the contemporaneous number of new firms in a block controlling for neighborhood-year fixed effects. These neighborhoods are the result of the algorithm process described in the text and they are city-industry specific.

Table 4: Main Results

(a) end-of-year Revenue

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log (Average Employment Same Industry$_{t-1}$)	0.120**	0.171***	0.137***	0.157**	0.143**	0.0921**	0.108**	0.120**
	(0.0523)	(0.0416)	(0.0229)	(0.0569)	(0.0515)	(0.0411)	(0.0418)	(0.0526)
Log (Average Revenue Same Industry$_{t-1}$)	0.0273**	0.0351*	0.0365**	0.0175**	0.0282**	-0.00245	0.0345**	0.0273**
	(0.0129)	(0.0187)	(0.0134)	(0.00817)	(0.0128)	(0.00744)	(0.0123)	(0.0127)
Log (# Incumbents Same Industry$_{t-1}$)	-0.114*	-0.00935	0.00371		-0.114**	0.0357	-0.00624	-0.114*
	(0.0599)	(0.0332)	(0.0305)		(0.0572)	(0.0412)	(0.0387)	(0.0602)
No Incumbent Economic Activity$_{t-1}$	0.108	0.299*	0.431**		0.116	-0.0586	0.226**	0.108
	(0.134)	(0.174)	(0.166)		(0.133)	(0.0663)	(0.115)	(0.137)
Previous Ownership Experience	0.0490***	0.0104***	0.0175***		0.0384***	0.102***	0.0481**	0.0490***
	(0.0121)	(0.00286)	(0.00395)		(0.00846)	(0.0157)	(0.0157)	(0.0120)
Previous Industry Experience	0.0732***	0.0672***	0.0674***		0.0389***	0.0606***	0.0617***	0.0732***
	(0.00700)	(0.00772)	(0.00502)		(0.00848)	(0.00657)	(0.00493)	(0.00688)
Number of Owners	-0.0697	-0.00356	-0.0492		-0.143**	0.0379	-0.0137	-0.0697
	(0.0586)	(0.0629)	(0.0397)		(0.0554)	(0.0760)	(0.0417)	(0.0604)
Corporate Partner	0.246	0.248	0.131		0.202*	0.549**	0.806***	0.246*
	(0.152)	(0.173)	(0.277)		(0.116)	(0.220)	(0.153)	(0.146)
Family Control	0.470**	0.482***	0.556***		0.631***	0.547***	0.744***	0.470**
	(0.149)	(0.120)	(0.0974)		(0.146)	(0.0821)	(0.0868)	(0.149)
Number of Observations (Startups)	30,318	41,112	81,319	30,318	37,944	81,586	142,465	30, 318

(b) end-of-year Employment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log (Average Employment Same Industry$_{t-1}$)	0.218***	0.333***	0.244***	0.252***	0.220***	0.209***	0.279***	0.218***
	(0.0442)	(0.0443)	(0.0221)	(0.0459)	(0.0439)	(0.0383)	(0.0269)	(0.0449)
Log (Average Revenue Same Industry$_{t-1}$)	-0.0257*	-0.0348**	-0.0238**	-0.0170**	-0.0259*	-0.0135**	-0.0212***	-0.0257*
	(0.0139)	(0.0131)	(0.00758)	(0.00776)	(0.0139)	(0.00613)	(0.00597)	(0.0136)
Log (# Incumbents Same Industry$_{t-1}$)	-0.0305	-0.117**	-0.00839		-0.0314	-0.00939	-0.0207	-0.0305
	(0.0613)	(0.0437)	(0.0225)		(0.0605)	(0.0286)	(0.0316)	(0.0622)
No Incumbent Economic Activity$_{t-1}$	-0.194	-0.233*	-0.145		-0.192	-0.0584	-0.130**	-0.194
	(0.171)	(0.134)	(0.0918)		(0.172)	(0.0677)	(0.0645)	(0.171)
Previous Ownership Experience	0.00138	0.00141	0.00716*		0.00214	0.0233**	0.00337	0.00138
	(0.00551)	(0.00234)	(0.00419)		(0.00605)	(0.0103)	(0.00533)	(0.00549)
Previous Industry Experience	0.0836***	0.0692***	0.0746***		0.0775***	0.0780***	0.0782***	0.0836***
	(0.00685)	(0.00704)	(0.00407)		(0.00628)	(0.00542)	(0.00405)	(0.00715)
Number of Owners	0.228***	0.177***	0.117***		0.200***	0.0842**	0.174***	0.228***
	(0.0521)	(0.0518)	(0.0295)		(0.0483)	(0.0358)	(0.0258)	(0.0543)
Corporate Partner	0.632***	0.565***	-0.0247		0.539***	0.931*	0.966***	0.632***
	(0.166)	(0.156)	(0.292)		(0.150)	(0.512)	(0.184)	(0.162)
Family Control	0.114	0.114	-0.0153		0.0995	-0.0844	-0.0202	0.114
	(0.181)	(0.153)	(0.100)		(0.150)	(0.113)	(0.0860)	(0.184)
Number of Observations (Startups)	20,675	41,570	66,678	20,675	25,088	60,740	109,411	20,675
Sample	Away	Away	Away	Away	Inst. Entrants	Home	Home/Away	Away
Neighborhood Year FE	YES	NO	FSA	YES	YES	YES	YES	NO
Cluster Level	NeighYear	Year	NeighYear	NeighYear	Neigh Year	NeighYear	NeighYear	Neigh

Notes: Table report results of estimating Equation (2). Dependent variable is the level of end-of-year revenue (Panel a) or end-of-year employment (Panel b) where end-of-year means end of the founding year of the startup. Controls are labeled in the first column. The variables of interest are given by $Log(X_{t-1})$ that correspond to an aggregation of characteristics (employment or revenue) of incumbent firms of industry j in block b within neighborhood n at time $t-1$. Equation is estimated using PPML. Standard errors are clustered at the neighborhood-year level unless stated otherwise.

Table 5: Local Spillovers over Survival and Future Location Choices

	(1) Alive $t+1$	(2) Alive $t+5$	(3) Move Blocks $t+1$	(4) Move Neighborhoods $t+1$	(5) Move Blocks $t+5$	(6) Move Neighborhoods $t+5$
Log (Average Employment Same Industry$_{t-1}$)	0.000245	0.0177**	-0.159***	-0.173***	-0.0338***	-0.0341***
	(0.00224)	(0.00625)	(0.0199)	(0.018)	(0.00479)	(0.00369)
Log (Average Revenue Same Industry$_{t-1}$)	0.000503	0.0000189	0.000953	0.00110	0.00367***	0.00369***
	(0.000632)	(0.00191)	(0.00479)	(0.00434)	(0.00133)	(0.00133)
Number of Observations (Startups)	29,621	28,989	19,714	24,855	28,528	37,529

Neighborhood-Year FE, Industry-Year FE and City-Year FE. Sample: Away. Controls include: number of incumbent firms, dummy of no economic activity, owners previous industry experience, owners previous entrepreneurship experience, total number of owners, dummy of corporate partner, and dummy of family control. Standard errors clustered at the neighborhood-year level.

Table 6: Spatial Decay

	Same Block	First Ring	Second Ring	Third Ring	Number of Observations (Startups)
a) end-of-year Revenue					
Log (Average Employment Same Industry$_{t-1}$)	0.118**	0.0621*	-0.0230	0.0154	30,318
	(0.0489)	(0.0350)	(0.0508)	(0.0280)	
Log (Average Revenue Same Industry$_{t-1}$)	0.0259**	0.00739	-0.0168*	-0.00138	
	(0.0123)	(0.0133)	(0.0101)	(0.00802)	
b) end-of-year Employment					
Log (Average Employment Same Industry$_{t-1}$)	0.212***	0.0280	0.0239	0.0118	20,675
	(0.0428)	(0.0332)	(0.0275)	(0.0242)	
Log (Average Revenue Same Industry$_{t-1}$)	-0.0255*	-0.0262***	-0.000222	0.00649	
	(0.0139)	(0.00641)	(0.0107)	(0.00765)	

Each panel is one regression. Coefficients correspond to a dummy for a group of industries interacted with the variable in the first column. The following set of fixed effects is included in each regression Neighborhood-Year FE, Industry-Year FE and City-Year FE. Sample: Away. Controls include: number of incumbent firms, dummy of no economic activity, owners previous industry experience, owners previous entrepreneurship experience, total number of owners, dummy of corporate partner, and dummy of family control. Standard errors clustered at the neighborhood-year level.

Table 7: Heterogeneity Across Industries

	Information and Financial Services	Manufacturing	Professional and Business Services	Retail, Leisure and Hospitality	Transport and Wholesale Trade	Number of Observations (Startups)
a) end-of-year Revenue						
Log (Average Employment Same Industry$_{t-1}$)	-0.227**	0.812***	0.0739	0.209***	0.340**	30,318
	(0.106)	(0.189)	(0.114)	(0.0459)	(0.126)	
Log (Average Revenue Same Industry$_{t-1}$)	0.0530**	-0.0802*	0.0293*	0.00145	0.0143	
	(0.0187)	(0.0442)	(0.0159)	(0.0129)	(0.0185)	
b) end-of-year Employment						
Log (Average Employment Same Industry$_{t-1}$)	0.000781	0.722***	0.200	0.202***	0.202*	20,675
	(0.234)	(0.175)	(0.131)	(0.0385)	(0.115)	
Log (Average Revenue Same Industry$_{t-1}$)	-0.00499	-0.0731**	-0.0370*	-0.0268*	-0.0102	
	(0.0283)	(0.0344)	(0.0215)	(0.0137)	(0.0200)	

Each panel is one regression. Coefficients correspond to a dummy for a group of industries interacted with the variable in the first column. The following set of fixed effects is included in each regression Neighborhood-Year FE, Industry-Year FE and City-Year FE. Sample: Away. Controls include: number of incumbent firms, dummy of no economic activity, owners previous industry experience, owners previous entrepreneurship experience, total number of owners, dummy of corporate partner, and dummy of family control. Standard errors clustered at the neighborhood-year level.

Table 8: Heterogeneity to Type of Exposure

	Baseline (same-industry)	All	Downstream	Upstream	Occ. Similarity
a) end-of-year Revenue					
Log (Average Employment$_{t-1}$)	0.120**	0.186***	0.0437**	0.0407**	0.141***
	(0.0523)	(0.0296)	(0.0147)	(0.0149)	(0.0261)
Log (Average Revenue$_{t-1}$)	0.0273**	0.0319**	0.0907***	0.100***	0.0637***
	(0.0129)	(0.0145)	(0.0130)	(0.0135)	(0.0131)
b) end-of-year Employment					
Log (Average Employment$_{t-1}$)	0.218***	0.209***	0.0359**	0.0379**	0.173***
	(0.0442)	(0.0331)	(0.0137)	(0.0137)	(0.0280)
Log (Average Revenue$_{t-1}$)	-0.0257*	-0.0101	0.0418**	0.0435**	0.000121
	(0.0139)	(0.0173)	(0.0202)	(0.0205)	(0.0199)

Each column is a different regression based in the specification of column (1) of Table (4), except the variables in the left are weighted by cross-industry weights. Sample: Away. Controls include: number of incumbent firms, dummy of no economic activity, owners previous industry experience, owners previous entrepreneurship experience, total number of owners, dummy of corporate partner, dummy of family control, Neighborhood-Year FE, Industry-Year FE and City-Year FE. Standard errors clustered at the neighborhood-year level.

Table 9: Descriptives of the Estimates Probabilities and Chosen Locations

Variable	**Mean**	**Std.Dev.**	**Min**	**Max**
Number of Neighborhoods	189.6	178.1	1.0	573
Rank of Chosen Neighborhood	31.29	67.48	1.0	573
Probability of Chosen Neighborhood	0.245	0.345	0.0	1.0
Maximum Neighborhood Probability	0.575	0.286	0.04	1.0
Distance of Chosen Neighborhood	168.8	686.5	0.0	4570
Probability of First Residential Neighborhood	0.691	0.461	0.0	1.0
Rank of First Residential Neighborhood	100.3	91.34	1.0	287

Note: These estimated probabilities are the result of a estimating the conditional choice model using entrepreneurs demographics and distance to their first residential neighborhood as instrument for neighborhood choices.

Table 10: Block Level Spillovers Across Neighborhoods

(a) End of Year Revenue

	Only Within Neighborhoods		Distance to First Residence		Distance to Prev Residence	
	$\delta^{(j)}_{nt}$	$\delta^{(j)}_{nt}$	$\delta^{(j)}_{nt} \times \mathrm{P}^{star(i,j)}_{nt}$	$\delta^{(j)}_{nt} \times \mathrm{P}^{home(i,j)}_{nt}$	$\delta^{(j)}_{nt} \times \mathrm{P}^{star(i,j)}_{nt}$	$\delta^{(j)}_{nt} \times \mathrm{P}^{home(i,j)}_{nt}$
Log (Average Employment Same Industry$_{t-1}$)	0.120**	0.169***	0.152***	0.176***	0.184***	0.172***
	(0.0523)	(0.0197)	(0.0323)	(0.0311)	(0.0389)	(0.0362)
Log (Average Revenue Same Industry$_{t-1}$)	0.0273**	0.0381***	0.0299**	0.0183**	0.0165**	0.0198**
	(0.0129)	(0.0100)	(0.0109)	(0.00699)	(0.00767)	(0.00909)
Number of Observations (Startups)	30,318	247,594	125,676	125,676	125,676	125,676

(b) End of Year Employment

	Only Within Neighborhoods		Distance to First Residence		Distance to Prev Residence	
	$\delta^{(j)}_{nt}$	$\delta^{(j)}_{nt}$	$\delta^{(j)}_{nt} \times \mathrm{P}^{star(i,j)}_{nt}$	$\delta^{(j)}_{nt} \times \mathrm{P}^{home(i,j)}_{nt}$	$\delta^{(j)}_{nt} \times \mathrm{P}^{star(i,j)}_{nt}$	$\delta^{(j)}_{nt} \times \mathrm{P}^{home(i,j)}_{nt}$
Log (Average Employment Same Industry$_{t-1}$)	0.218***	0.316***	0.319***	0.313***	0.312***	0.294***
	(0.0442)	(0.0167)	(0.0358)	(0.0396)	(0.0428)	(0.0400)
Log (Average Revenue Same Industry$_{t-1}$)	-0.0257*	-0.0159**	-0.0260**	-0.0223**	-0.0310**	-0.0229**
	(0.0139)	(0.00532)	(0.00919)	(0.0100)	(0.00986)	(0.0111)
Number of Observations (Startups)	20,675	249,938	88,687	87,760	89,666	86,325
Fixed Effects						

FE: $\delta^{(j)}_{nt} \times \mathrm{P}^{(i,j)}_{nt}$, Industry-Year FE and City-Year FE.
Same controls as benchmark regression.