The Concentration of Economic Activity Within Cities: Evidence from New Commercial Buildings

Thiago Patto[*]

July 22, 2023

Work in progress draft. Please do not circulate.

Abstract

I explore the opening of large commercial buildings to study how economic activity is affected locally by a surge in employment density. I combine property tax records and linked employer-employee data to identify newly opened buildings that received many workers and examine the effects on nearby neighborhoods. My methodology combines the typical ring approach (comparing locations nearer vs. farther) with a procedure that explores variation within neighborhoods with a similar probability of observing a new building in its vicinity. I find that neighborhoods within 250 meters of a new building experience an 11.1% differential increase in employment, driven by high-skilled offices and local services. I estimate that for every two additional jobs created by high-skilled offices, one job is created by local services. I also show suggestive evidence that the productivity of high-skilled offices is affected. There is an increase of 3.4 percentage points in the share of college-educated workers and a 6.1% increase in wages for this sector, which seems to be driven to a good extent by new firms. Overall, my findings indicate that both productivity spillovers and local demand are relevant elements of urban concentration. New buildings increase the productivity of high-skilled offices nearby, attracting more firms in this sector and raising the demand for non-tradable goods provided by local services. This sector, in turn, increases its presence as well.

[*]Insper, São Paulo, *thiagospms@al.insper.edu.br*

1 Introduction

The uneven distribution of economic activity within cities is readily perceptible and can be more pronounced than that between cities. As an example, in 2021, the New York Metropolitan Area accounted for about 8% of the United States' GDP, with Manhattan alone responsible for 40% of this contribution despite covering only 0.3% of the land area. This trend is prevalent in numerous major cities worldwide, where companies flock to the most attractive neighborhoods despite the congestion costs involved, particularly the elevated rental prices.

There is a conventional view among economists that agglomeration forces are an essential element in understanding this phenomenon. However, while the literature has primarily focused on workers' location decisions, fewer studies have approached this issue from the firm's perspective.[1] This matter is especially relevant if we consider that different firms might be affected by density to varying degrees and for different reasons. For example, while some industries are drawn to high-employment locations by the greater demand for their goods, others could benefit from productivity spillovers. Moreover, if highly productive firms are more sensitive to the benefits of agglomeration, as some papers have conjectured, they will tend to sort into high-density neighborhoods.[2] These topics have been present in studies at the regional level but are yet to be explored within urban areas.

Using the opening of new large commercial buildings in Sao Paulo as a natural experiment, this article helps to fill this gap by studying how a sudden increase in urban concentration in one location impacts neighborhoods in the vicinity. The analysis focuses on how different sectors respond to these shocks and compares neighborhoods more or less exposed to these buildings to identify the effects. My sector classification aims to differentiate between establishments that produce non-tradable goods and therefore are more dependent on local demand (local services) and those that produce tradable goods (high- and low-skilled offices).[3]

With 21.7 million inhabitants, Sao Paulo is one of the ten largest metropolitan areas globally and the largest in America. It has almost doubled its population in 50 years, thus providing a compelling setting for the study.[4] Furthermore, it contains high-quality data with detailed location information. I use two administrative sources that can be accurately geocoded to perform the empirical analysis: property tax records from the Sao Paulo municipal government (IPTU) and matched employer-employee data from the

[1] This trend has arisen greatly due to the popularity of quantitative spatial models, which consider individuals who choose where to live and work but abstract from employers' location decisions. Some examples are Ahlfeldt et al. (2015), exploring the division and reunification of Berlin; Heblich et al. (2020) and Tsivanidis (2022), exploring new transit infrastructures. Redding (2022) surveys this literature. A few exceptions focusing on firms' location include Arzaghi and Henderson (2008) and Baum-Snow et al. (2021).

[2] See, for instance, Combes et al. (2012) and Gaubert (2018).

[3] The words "firm" and "establishment" are used interchangeably throughout the text, and I use the latter to emphasize a physical unit. The sector classification addresses the issue of multi-establishment firms. Section 2.2 discusses the matter, and Appendix A gives more details.

[4] Source: www.oecd-ilibrary.org//sites/9b73e35d-en/index.html?itemId=/content/component/9b73e35d-en#

Ministry of Labor (RAIS). I use them for two purposes. First, I combine the datasets to identify buildings that were recently inaugurated and received a significant number of workers. This procedure allows me to interpret these constructions as employment shocks, in the spirit of Greenstone et al. (2010) exploring the opening of large plants, but at a more local level. Secondly, I use them to characterize economic activity in neighborhoods, defined as a contiguous set of 200-meter square cells.

From a sample of 43 new commercial buildings, I develop a difference-in-differences model with staggered treatment adoption. I define treated and control neighborhoods based on the nearest building site and the year of treatment based on the first building within the range that characterizes that neighborhood as a treated unit. These definitions are necessary since some neighborhoods are close to more than one building.[5]

A central challenge in this empirical approach is that developers endogenously choose buildings' locations. I restrict the sample to cells within 1 km of a new building and explore small variations in distance to attenuate concerns about selection, as treated and control neighborhoods belong to the same regions and are presumably similar in several dimensions. I also take advantage of the panel structure of the data and estimate event-study specifications to check for pre-trends.

Nevertheless, it is still possible that neighborhoods face local shocks correlated with the likelihood of being close to a new building. To address this concern, I build on Qian and Tan (2021) and develop a procedure to compare neighborhoods with a similar probability of being treated. I estimate a propensity score model that uses neighborhood characteristics prior to buildings' inaugurations to predict which locations are more likely to receive a new construction. From the predicted values, I create a measure that expresses the ex-ante probability that a cell observes a new building in its vicinity. Then, based on this measure, I separate treated and control neighborhoods into terciles and estimate treatment effects exploring variation only within these groups.

The results show significant effects on local economic activity. Neighborhoods within 250 meters of a new building experience an 11.1% differential increase in employment compared to neighborhoods between 500 meters and 1 km. Local services and high-skilled offices account for virtually all of this growth, with differential increases of 13.4% and 20.9%, respectively. Considering the average employment one period prior to the treatment, it means that approximately one job is created in local services for each additional two jobs created in high-skilled offices.

I also show evidence consistent with the idea that the productivity of high-skilled offices is affected in treated neighborhoods. There is a differential increase of 3.4 percentage points in the share of college-educated workers and a differential increase of 6.1% in the average wage premium paid by this sector.[6]

[5] Neighborhoods that receive a new building are discarded from the estimations.

[6] I construct the average wage premium from establishment fixed effects estimated separately for each year from wage regressions.

However, additional evidence suggests that these effects are driven to a good extent by changes in firm composition, meaning that the sorting of high-wage firms in treated neighborhoods may be relevant to understanding the findings.

To interpret the results, I propose a stylized model of firm location choice built on Ahlfeldt et al. (2015) that yields an equilibrium distribution across neighborhoods. In the model, firms belong to different sectors and produce goods whose prices are defined by the larger economy, with the exception of local services, whose price is defined at the neighborhood level. Each pair neighborhood-sector has a local TFP, treated in principle as fixed. Because I assume that workers in a given neighborhood spend a fraction of their income on local services, this sector responds to changes in local economic activity driven by other sectors. I demonstrate that when there is a shift in the TFP of one sector in one neighborhood, this sector expands its presence in the neighborhood, leading to an increase in wages and employment.[7] Local services are then indirectly benefited and also expand. Thus, the estimated growth ratio of 0.5 between local services and offices can be interpreted as a multiplier effect.

I also explore the impacts of new buildings in other dimensions. I show that local services expand at the extensive margin in treated neighborhoods, with a 7.7% differential increase in the number of establishments. Event study point estimates suggest that offices are somewhat affected as well. Regarding the supply of floor space, I find no evidence of an impact on treated neighborhoods, suggesting that the increase in the number of establishments occurs through a decrease in vacancy rates.

Taken together, these findings indicate that productivity spillovers and local demand are fundamental drivers of spatial concentration. When a new commercial building opens, local employment increases rapidly and affects the productivity of nearby neighborhoods for high-skilled offices, especially high-wage firms. Because of that, this sector increases its activities in these locations. This process, in turn, raises the demand for local services and results in a higher presence of this sector as well.

This paper adds to the literature on the distribution of economic activity within cities, particularly on local agglomeration forces, pioneered by Arzaghi and Henderson (2008). Examples include cross-sectional studies (Rosenthal and Strange, 2020; Liu et al., 2020, 2022) and studies that rely on quantitative spatial models to estimate agglomeration effects (Ahlfeldt et al., 2015; Heblich et al., 2020; Tsivanidis, 2022). I go further than this literature by developing a new methodology that explores local shocks to provide evidence consistent with the existence of productivity spillovers at the neighborhood level.

In this sense, my findings are closely related and complementary to Baum-Snow et al. (2021). They estimate productivity externalities at a similar scale but using a peer effects model and firm revenue as a measure of productivity. While their approach delivers the estimation of structural parameters used to run

[7] In the model, firms face an upward-sloping labor supply curve, which generates a connection between local TFP and wages.

counterfactuals, it restricts the analysis to high-skilled services. This paper, on the other hand, provides a broader description of the process of urban concentration.

More generally, this paper also relates to studies on agglomeration economies (see Moretti, 2011; Combes and Gobillon, 2015, for related surveys). This literature has emphasized different aspects that influence the composition of cities, such as the sorting of firms (Combes et al., 2012; Gaubert, 2018), the urban wage premium and the sorting of workers (Combes et al., 2008; Baum-Snow and Pavan, 2012; De La Roca and Puga, 2017) and the interplay between tradable and non-tradable sectors (Moretti, 2010; Faber and Gaubert, 2019). My contribution is to approach some of these topics from a within-city perspective.

Finally, there is extensive literature that explores spatially distributed treatment effects. Two strands that are of particular interest are studies that explore the entry of large firms (Greenstone et al., 2010; Qian and Tan, 2021) and the construction of new residential buildings (Asquith et al., 2021; Pennington, 2021). My empirical analysis combines and adapts insights from these studies and provides a new approach that addresses potential endogeneity issues.

The remainder of the paper is structured as follows. Section 2 describes the data and provides some stylized facts that speak to the motivation of this study. Section 3 presents the conceptual framework, and Section 4 details the empirical approach. Section 5 presents the results, and Section 6 provides the robustness checks. Finally, Section 7 concludes.

2 Data and Descriptive Evidence

2.1 Data Sources

RAIS covers Brazil's formal labor market, with very few exceptions. It contains information about establishments at the annual level - with invariant identifiers for both the establishment and the associated firm - and information about workers at the job record level - with identifiers for the individual and the associated establishment. Data available at the individual level include educational attainment, tenure, occupation, earnings and weekly contracted hours. I use the latter two to compute monthly wages. An essential piece of information for this study is the establishments' complete addresses, which employers report annually.

IPTU is an annual panel that contains all formal real estate in the municipality of Sao Paulo at the unit level.[8] It contains information related to tax collection, such as properties' purpose (commercial/residential), terrain and construction area, construction year and number of floors.

[8] A building is typically a collection of multiple units (apartments or offices), but in some cases, it can be a single observation in the data if there is only one landlord for the entire building.

I use RAIS to create a panel of private sector establishments in the Sao Paulo Metropolitan Area between 2003 and 2017. Based on this sample, I also construct an annual panel of individuals linked to these workplaces and use these data to estimate separately, for each year, wage premiums for each establishment according to the following expression[9]

$$\log w_{it} = X_{it}\Gamma_t + \psi_{j(i)t} + \nu_{it} \quad , \tag{1}$$

where $\log w_{it}$ is the log wage of individual i in year t, X_{it} is a group of controls, $\psi_{j(i)t}$ is an indicator for establishment j where individual i works in year t, and ν_{it} is an error term.[10] The estimated $\hat{\psi_{j(i)t}}$ represent the wage premia used in the analysis.

Then, observations in RAIS and IPTU are geocoded, with a rate of success of about 97% and 99%, respectively. More than 80% of the addresses of both RAIS and IPTU were successfully geocoded without imputation, i.e., with very high precision. These rates do not vary significantly across years.

The analysis is performed at the neighborhood level, defined as a 200-meter square cell. For this purpose, I split Sao Paulo's territory into a contiguous set of cells and use the successfully geocoded observations to compute the variables of interest. While IPTU gives information about floor space supply, RAIS provides a detailed description of economic activity.

2.2 Sector Classification

Another key piece of information in RAIS is the 5-digit industry code (CNAE) used to classify establishments into four groups. In general terms, the classification can be described as follows:[11,12]

- Local services: includes retail, food, bank agencies, gyms, personal care and maintenance services
- High-skilled offices: includes information and communication, professional services and finance (except retail banking)
- Low-skilled offices: includes administrative, support and health
- Non-offices: includes manufacturing, wholesale, education, utilities and transportation

Local services encompass those establishments that produce non-tradable goods whose demand is primarily local. They benefit from being geographically close to potential consumers. The second and third groups

[9] If a worker has more than one job record in the same year, I keep the one with the highest tenure (or higher wage if there is a tie).

[10] Control variables include a cubic polynomial in age fully interacted with gender and college indicators, a cubic polynomial in tenure interacted with a college indicator and 4-digit occupation fixed-effects.

[11] The detailed correspondence between 5-digit CNAE and sectors is available upon request.

[12] There is a small group of establishments whose industry code changes over time. I deal with this by creating a fixed classification using the most observed value. In general, these changes occur between similar industries.

represent establishments from industries typically found in urban areas and produce goods and services relatively more tradable than local services. The term "office" is used since they are the main consumers of floor space in areas of high employment density. High-skilled offices refer to industries that offer highly tradable specialized services, ranging from city to regional level. These industries have received particular attention in the literature since their presence in big cities is especially pronounced and has been increasing in recent years (Davis and Dingel, 2020; Davis et al., 2020; Eckert et al., 2020). Low-skilled offices, in turn, are industries that produce less specialized services and whose demand expands beyond neighborhoods but not much further. Finally, the last group refers to establishments in the private sector that do not belong to either of the three sectors mentioned above. It contains industries that produce tradable or partially tradable goods that tend to locate in less dense areas. Despite being a highly heterogeneous group, its inclusion in the analysis allows me to investigate to what extent changes in a neighborhood are due to an increase in aggregate density or a substitution effect across sectors.[13]

To guarantee an accurate classification, I also consider the issue of multi-establishment firms. In many cases, the sectoral classification is properly done by firms at the establishment level, which means that different establishments from the same firm can have different CNAE codes depending on their purpose. However, there are some exceptions of a single CNAE code for the entire firm, which can lead to mismeasurements, such as labeling administrative establishments of a retail chain as a local service. To deal with that, I develop a more complex classification for firms with more than 20 establishments in the Sao Paulo Metropolitan Area that explores the occupation composition of establishments. Details are provided in Appendix A.[14]

2.3 Descriptive Evidence

In this section, I document two facts about the distribution of economic activity in Sao Paulo that speak to the motivation of this paper: a) some sectors are more concentrated than others, and b) more productive firms tend to be more concentrated. I focus on the city's districts where the empirical analysis is performed.

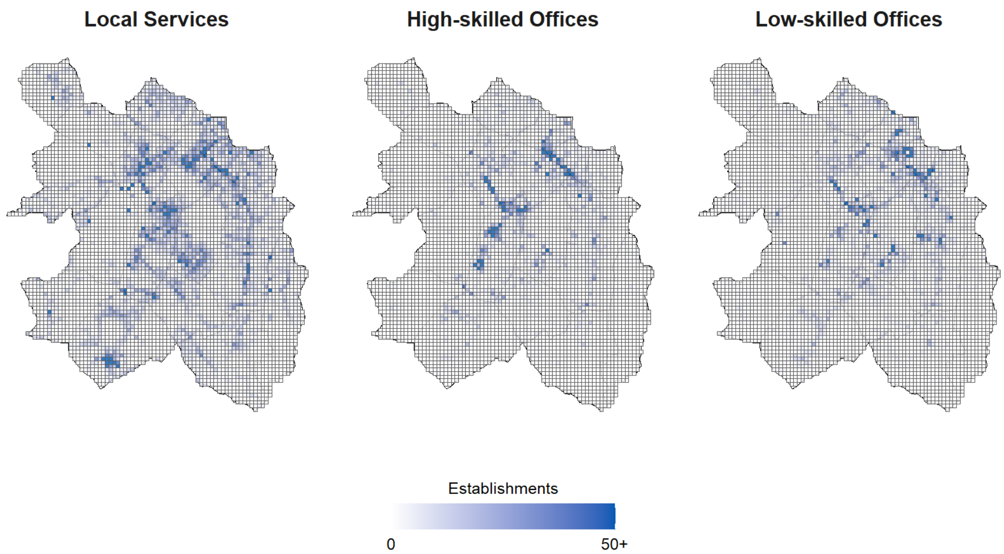
Related to the first point, Figure 1 shows the spatial distribution of establishments classified as local services, high- and low-skilled offices (as defined in the previous section) in 2010. The blue areas indicate where establishments are located.

There are significant differences in the concentration level of each of these sectors. Local services spread relatively more across the city, but offices exhibit a more concentrated pattern, especially high-skilled offices. Most of the agglomeration takes place around three important avenues of Sao Paulo: Paulista (on the

[13]The only establishments excluded from the analysis are those in the public sector.

[14]Because this procedure needs a reasonable number of establishments to work, some will remain misclassified. This situation is more common in specific industries, such as manufacturing and agriculture, where some establishments are likely to be typical offices. Nonetheless, I choose to be conservative and label these establishments as non-offices.

Figure 1: Spatial Distribution of Employment by Sector



Notes: The figure displays, for 2010 , the spatial distribution of establishments by cell for different sectors.

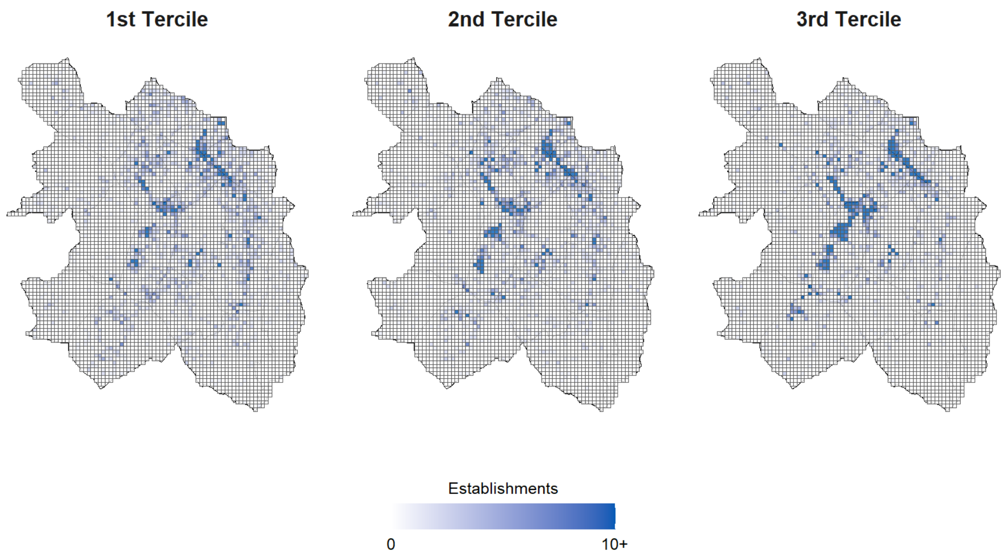
northeast side of the map), Brigadeiro Faria Lima and Luis Carlos Berrini (at the center of the map). They represent, respectively, the old and new business centers of Sao Paulo and contain the highest employment densities in the city.[15]

One potential factor behind this pattern is related to the nature of the goods produced by these sectors. Businesses like restaurants and retail require proximity to potential customers, so they prefer to locate where people live or circulate. Thus, even though high-density neighborhoods tend to have a larger provision of local services, firms in this sector also have some incentives to locate in low-density neighborhoods. On the other hand, firms in the financial or tech sector do not rely on local demand but may benefit from productivity spillovers in high-density locations. Depending on the magnitude of these spillovers, it can discourage firms from locating in isolated neighborhoods.

Regarding the second point, Figure 2 displays, for 2010, the distribution of establishments in high-skilled offices for different terciles of establishment wage premium. It is worth noting that establishments in the first tercile are more dispersed than in the third tercile, i.e., the concentration level increases with the wage premium. To show that this fact is not entirely driven by industry composition, in Figure C1 I present the

[15] Historically, the first employment boom of Sao Paulo occurred in the region known as the historical center, located just north of the region depicted in Figure 1. Then, in the 1950s, employment began to move gradually to Avenida Paulista. The traditional mansions were demolished to clear space for commercial buildings, which hosted the headquarters of many large companies. Avenida Brigadeiro Faria Lima began to gain relevance in the 1970s, and by the 1990s, it was already one of the densest employment areas. Nowadays, it is considered the most important business center of Sao Paulo, especially for industries such as finance and tech. Employment in Avenida Luis Carlos Berrini followed the same trend as Faria Lima but with a few years of delay

Figure 2: Spatial Distribution of High-skilled Offices by Wage Premium Tercile



Notes: The figure displays, for 2010, the spatial distribution of high-skilled offices by cell for different terciles of establishment wage premium. For more details about the estimation of these premia, see Section 2.1.

same analysis focusing on one specific industry, namely finance, with qualitatively similar results. These figures suggest that more productive firms have a stronger preference for concentration, possibly because their productivity is more sensitive to agglomeration effects.

These pieces of evidence underline that the preference to locate in high-density neighborhoods is heterogeneous and suggest that agglomeration may affect firms differently. In light of this, I now propose a theory of firm location choice that rationalizes urban concentration as a combination of spatial differences in productivity and local demand effects. Additionally, I discuss how firm heterogeneity can be relevant in this context.

3 Theoretical Framework

The first relevant question to study the effects of new commercial buildings on local economic activity is how this influence can be exerted. I consider three possible channels. First, new buildings can represent a shift in local demand as workers and firms that occupy the newly created space may spend some income on locally produced goods. Secondly, new buildings may impact local amenities in various ways, from landscape to safety, including indirect effects from the government responding to these changes (e.g., investments in public infrastructure). Finally, the surge in employment may impact local productivity through spillover effects, which can occur, for instance, via local diffusion of knowledge or improvements in the quality of firm-worker matching.

In this section, I present a stylized spatial model of firm location choice built on Ahlfeldt et al. (2015) that contains elements of these three features. I focus my analysis on changes in local productivity, as I argue that this particular channel is key to understanding the process of urban concentration. I consider a variation in local TFP for a specific sector to obtain testable predictions and compare them with the empirical evidence.

In order to connect the model with the results, I assume that new buildings are unlikely to represent a significant departure from the previous equilibrium of the city but have sizable local effects. This assumption allows me to treat some variables as fixed and obtain predictions from simple comparative statics. All derivations are detailed in Appendix B.

3.1 A Model of Firm Location Choice

Consider a city with multiple discrete neighborhoods indexed by n and a continuum of firms in discrete sectors $\overline{E}_s$ choosing where to locate.[16] Firms are indexed by e. Each pair sector-neighborhood (s, n) has a local TFP $A_{s,n}$, which is assumed to be given for now. In order to maximize profits, firms first choose their location and then the amount of labor ℓ and floor space f to produce a a homogeneous good within sectors. While rent prices r_n are taken as given, each firm faces an upward-sloping labor supply curve and hence needs to choose the optimal wage $w_{e,s,n}$. To make it simple, I assume that the labor supply curve is identical for all firms in the same pair (s, n), and since they solve the same profit maximization problem, in practice, wages will be sector-neighborhood specific, i.e., $w_{e,s,n} = w_{s,n}\ \forall e$.

There are two groups of sectors in this economy, offices and local services (LS), which differ basically in terms of the range of their markets. Offices sell their goods all over the city and to the broader economy at a common price p_s, treated as fixed. On the other hand, local services sell their goods locally at a price $p_{LS,n}$. For simplicity, I assume that this good is entirely consumed by workers in the same neighborhood, using a fraction δ of their income.

If a firm chooses to locate in neighborhood n, it will choose wages and the amount of floor space according to

$$\pi_{e,s,n} = \theta_{e,n} \cdot \begin{cases} \max\limits_{f,w} \left[p_s A_{s,n} f^{\beta} \ell_n^{1-\beta}(w) - r_n f - w\ell_n(w)\right] \text{ if } s \neq LS \\ \max\limits_{f,w} \left[p_{LS,n} f^{\beta} \ell_n^{1-\beta}(w) - r_n f - w\ell_n(w)\right] \text{ if } s = LS, \end{cases} \tag{2}$$

where $\beta < 1$ is the share of expenditures in floor space and $\theta_{e,n}$ is a preference shock that firms draw

[16] The choice of a closed city is justified for convenience since it facilitates the derivation of the results, but the model can be easily modified to accommodate an open city version in which the expected profit for each sector must equal a common reservation level of profit.

independently from a Frechét distribution with cdf $F(\theta) = e^{-\theta^{-\eta}}$. It represents idiosyncratic preferences entrepreneurs would have for specific locations, e.g., being closer to where they live. Note that the TFP of local services is assumed to be the same in all neighborhoods ($A_{LS,n} = 1 \ \forall n$).

Firms solve this problem considering their supply curve in n, given by

$$\ell_n = B_n \left(\frac{w}{p_{LS,n}^{\delta}} \right)^{\varepsilon}. \tag{3}$$

In this expression, ε represents the elasticity of labor with respect to wages, and B_n is the firm commuter market access (FCMA) (Tsivanidis, 2022). This term captures how easy it is for firms in a neighborhood to attract workers. Appendix B shows a possible micro foundation for this expression, where individuals live in fixed neighborhoods and have idiosyncratic preferences across firms. Moreover, they value some neighborhoods more than others and take into account firms' location when choosing where to work.

Using the first order conditions of (2), the wage set by a firm in sector s that locates in n is given by

$$w_{s,n} = \left[\frac{1-\beta}{\beta} \frac{\varepsilon}{\varepsilon+1} \right] \beta^{1/(1-\beta)} \cdot \begin{cases} p_s^{\frac{1}{1-\beta}} \left(\frac{A_{s.n}}{r_n^{\beta}} \right)^{\frac{1}{1-\beta}} & \text{if } s \neq LS \\ \left(\frac{p_{LS,n}}{r_n^{\beta}} \right)^{\frac{1}{1-\beta}} & \text{if } s = LS \end{cases}. \tag{4}$$

Note that because the labor supply curve is upward sloping, the wages paid by offices are increasing in local TFP.

Firms choose their location by choosing the neighborhood that gives the highest profit. By combining the first-order conditions of (2) with the distribution of θ, it is possible to derive an expression that gives the number of firms in office sector s that chooses to locate in neighborhood n:

$$E_{s,n} = \left(\frac{A_{s,n}^{\chi} B_n}{r_n^{\chi\beta} p_{LS,n}^{\delta\varepsilon}} \right)^{\eta} \frac{\overline{E}_s}{\Phi_s}, \tag{5}$$

where $\Phi_s = \sum_i (A_{s,i}^{\chi} B_i / r_i^{\chi\beta} p_{LS,i}^{\delta\varepsilon})^{\eta}$ and $\chi = (1+\varepsilon)/(1-\beta)$

This expression tells us that the relative sectoral presence in a neighborhood depends positively on local TFP and FCMA, and negatively on rent and local services prices. Naturally, higher benefits will be counterbalanced in equilibrium by higher costs. However, it is worth noting that while prices faced by firms in a given neighborhood are the same, TFP can vary by sector and thus generate differences in their spatial distribution.[17]

For local services, the same procedure yields

[17] One could also consider differences in technology as another driver to rationalize variations in spatial distribution by sector, which in this model is represented by the common parameter β. However, while this channel could explain why some sectors are more likely to locate in low-rent price areas, it cannot account for differences in the concentration level of different sectors.

$$E_{LS,n} = \left(\frac{p_{LS,n}^{\chi-\delta\varepsilon} B_n}{r_n^{\chi\beta}}\right)^{\eta} \frac{\overline{E}_{LS}}{\Phi_{LS}}, \tag{6}$$

where $\Phi_{LS} = \sum_i (p_{LS,i}^{\chi-\delta\varepsilon} B_i / r_i^{\chi\beta})^{\eta}$. In this case, a higher price of local services increases the presence of firms in this sector since it positively affects their profits.

The equilibrium of this economy is characterized by the vector of prices r_n and $p_{LS,n}$ that solves two market clearing conditions. The first one equates supply and demand for local services in each neighborhood:

$$Y_{LS,n} E_{LS,n} = \frac{\delta}{p_{LS,n}} \sum_s w_{s,n} \ell_{s,n} E_{s,n}, \tag{7}$$

where $Y_{LS,n}$ is the amount of non-tradable goods produced in neighborhood n. Using (5), (6), and the first order conditions of (2) to solve for $Y_{LS,n}$, $w_{s,n}$ and $\ell_{s,n}$, it is possible to derive an expression for $p_{LS,n}$:

$$p_{LS,n} = \left[\left(\frac{1}{\chi/\delta\varepsilon - 1}\right) \frac{\Phi_{LS}}{\overline{E}_{LS}} \sum_{s \neq LS} \frac{\overline{E}_s p_s^{\chi}}{\Phi_s} \cdot A_{s,n}^{\chi(1+\eta)}\right]^{\frac{1}{\chi(1+\eta)}}. \tag{8}$$

Equation (8) relates $p_{LS,n}$ and $A_{s,n}$, showing that the price of local services is higher in more productive neighborhoods.

The second market clearing condition equates supply and demand for floor space in each neighborhood. Here, I assume that there is a competitive construction sector that uses land and capital as inputs. Land available in neighborhood n is fixed at $\overline{T}_n$. Assuming a Cobb-Douglas production function in which the share of expenditures on land is α, the floor space supply curve in neighborhood n can be written as $\frac{\overline{T}_n}{\alpha} r_n^{\frac{1-\alpha}{\alpha}}$. Equating this expression with floor space demand yields[18]

$$\frac{\overline{T}_n}{\alpha} r_n^{\frac{1-\alpha}{\alpha}} = \sum_s f_{s,n} E_{s,n}. \tag{9}$$

To derive an expression for r_n, I use again (5), (6) and the first order conditions of (2), together with Equation (8), to get

$$r_n = \left[K \frac{\overline{E}_{LS}}{\Phi_{LS}} \cdot \frac{B_n^{1+\eta}}{\overline{T}_n} \cdot p_{LS,n}^{(\chi-\delta\varepsilon)(1+\eta)}\right]^{\frac{1}{1/\alpha+\beta\chi(1+\eta)}}, \tag{10}$$

where $K = \frac{\alpha}{\delta}\left(\frac{\chi}{\varepsilon}\right)^{1-\varepsilon} \beta^{\chi-\varepsilon}$. Since $\chi - \delta\varepsilon > 0$, it is clear from Equations (8) and (10) that higher

[18] A more straightforward way to model the market for floor space is to assume a fixed supply, which could make sense since the analysis focuses on short-term impacts. In Section 5.1, I show that this quantity is not differentially affected in treated neighborhoods.

productivity is associated with higher prices of local services and floor space, as expected.

In order to discuss how this model can describe the spatial distribution of economic activity, I consider an increase in the TFP of one sector in one specific neighborhood. If the number of neighborhoods is sufficiently large, it is possible to treat Φ_s as constant for all sectors. This approximation allows for a comparative statics exercise that I summarize in two propositions.

Proposition 1 *Consider an increase in $A_{s,n}$. Assuming that $\dfrac{\partial \Phi_{s'}}{\partial A_{s,n}} \approx 0 \; \forall \; s'$, the number of firms of sector s in neighborhood n increases, whereas it decreases for all other office sectors. The number of firms providing local services also increases.*

Proof. See Appendix B ■

Having established from Equations (8) and (10) that the price of local services and floor space increase due to an increase in the local TFP of one sector, it is straightforward from Equation (5) that the presence of all office sectors $s' \neq s$ in neighborhood n are negatively affected. However, for sector s, the benefits of higher productivity exceed the increase in costs, and more firms will choose to locate there. Local services are also positively impacted due to higher demand for their goods.

Thus, there is an increase in the concentration of firms of sector s in neighborhood n that occurs due to a combination of heterogeneity in local productivity across sectors and common inputs, namely floor space and labor, whose price is locally defined. These inputs work as congestion forces that" select" the sectors with more presence in the neighborhood.

Proposition 2 *Consider an increase in $A_{s,n}$. Assuming that $\dfrac{\partial \Phi_{s'}}{\partial A_{s,n}} \approx 0 \; \forall \; s'$, nominal wages in sector s and local services increase, whereas it decreases for all other office sectors. Real wages (and employment) in sector s and local services tend to increase as well if η is not too high and, particularly for sector s if the elasticity of $p_{LS,n}$ with respect to $A_{s,n}$ is not close to 1.*

Proof. See Appendix B ■

The impact on nominal wages comes from the imperfectly competitive nature of the labor market. In this environment, equilibrium wages will be proportional to local TFP. If there is a productivity shock, profit maximization requires firms to raise wages. For local services, wages increase because they are also proportional to the output price, which is positively affected when one sector observes an increase in its productivity.

Interestingly, a positive shock in $A_{s,n}$ does not necessarily imply that $w_{s,n}/p^{\delta}_{LS,n}$ or $w_{LS,n}/p^{\delta}_{LS,n}$ will be higher. The intuition is that if η is large enough, i.e., if the dispersion of idiosyncratic preferences is low, rent prices will be more sensitive to changes in the price of local services. Because real wage set by firms depends negatively on rent prices, the net effect can be negative in some specific scenarios. For sector s, there is also the issue of how sensitive is $p_{LS,n}$ with changes in $A_{s,n}$, which makes the possibility of a negative impact even more unlikely. Since employment is a function of real wages, the same conclusion applies to this variable. Appendix B discusses this matter in more detail.

3.2 Spatial Sorting

I now consider the possibility that firms are ex-ante heterogeneous in productivity. Denote $A_{e,s,n}(\varphi_e, A_{s,n})$ the TFP of firm e in neighborhood n, which is a function of its own productivity φ_e and the sector-neighborhood productivity $A_{s,n}$.

In this new scenario, the probability $Pr_{e,s,n}$ that a firm e in an office sector s chooses to locate in neighborhood n is

$$Pr_{e,s,n} = \frac{1}{\Phi_{e,s}} \left(\frac{A^{\chi}_{e,s,n}(\varphi_e, A_{s,n}) B_n}{r_n^{\beta} p_{LS,n}^{\delta\varepsilon}} \right)^{\eta} , \tag{11}$$

where $\Phi_{e,s} = \sum_i (A^{\chi}_{e,s,i} B_i / r_i^{\chi\beta} p_{LS,i}^{\delta\varepsilon})^{\eta}$.

Following Gaubert (2018), an important case is when $A_{e,s,n}$ is log-supermodular in firm and sector-local productivity, which means that the local TFP of more productive firms increases disproportionately with A_{sn}. This assumption can be formally expressed by the condition $\dfrac{\partial^2 A_{esn}(\varphi_e, A_{sn})}{\partial \varphi_e \partial A_{sn}} > 0$.

In this situation, it is clear from (11) that firms with higher φ_e are more likely to choose highly productive neighborhoods. Furthermore, spatial sorting tends to increase with sectoral concentration. As certain neighborhoods become more productive for particular sectors, not only do these sectors expand their presence but average firm productivity increases as well.

4 Empirical Strategy

This section outlines the empirical approach of this paper, which explores the inauguration of new commercial buildings to construct a difference-in-differences model with staggered treatment adoption. In order to obtain estimates with a causal interpretation, a key identification challenge is to distinguish the specific impact of a new building from the more general causes that may have attracted the building to a particular site in the first place, given that developers endogenously choose where to construct. From a broader perspective, new constructions are essentially part of the gradual development of a city.

However, following Asquith et al. (2021) and Pennington (2021), I argue that there are variations that can be considered quasi-random at a more local level. For instance, after choosing an area of interest, developers choose the exact location among a few sites where the construction is feasible. The timing of the inauguration has an idiosyncratic component as well since the construction process is long and can be affected by issues not entirely controlled by developers, such as building permitting. Moreover, the size of the building may enhance these exogenous factors. Large constructions bring more complexity to the project and are more likely to face constraints related to geography and municipal legislation, which further increases uncertainty regarding the timing and location of the inauguration.

My methodology aligns with these arguments by focusing on short distances and intervals of time but refines this approach. I use a propensity score model to predict which cells are more likely to observe a new construction in its vicinity and use this information as a control variable in my specification. By doing so, I estimate the effects by exploring variation within groups of neighborhoods equally likely to have a new building nearby.[19] The panel structure is also a key element in my analysis. Observing neighborhoods before and after an inauguration makes it possible to difference out invariant factors that influence local economic activity. It also allows me to employ event-study regressions to verify the existence of pre-trends.

4.1 Selecting "Treatment" Buildings

From the IPTU data, I select an initial sample of new commercial properties with at least five floors. The restriction on the number of floors is because these new developments typically substitute low-density constructions or empty terrains. Moreover, I avoid dealing with lower buildings because sometimes they undergo re-classifications, making it difficult to observe them consistently in the data across the years. I also discard buildings with specific purposes that do not fit into this context, such as hotels, schools and religious temples.

This sample is not suitable on its own to identify shocks in local employment for two reasons. First, they do not provide good information about the buildings' inauguration year.[20] Secondly, some buildings may have a negligible employment impact if firms do not occupy them.

I overcome these limitations by merging RAIS and IPTU data using the address information. This procedure is helpful for many reasons. It reveals the increase in employment caused by each building and its composition. Furthermore, it allows me to observe the evolution of employment in each building and have a clear picture of when it starts to be populated.

[19] Although the propensity score is meant to address an endogeneity problem, it also touches on the issue of non-random exposure to exogenous shocks discussed by Borusyak et al. (2021). Since my measure of interest is the average propensity score within a certain radius of a neighborhood, those located in central areas of Sao Paulo tend to exhibit higher values. Thus, my specification also controls on some level for differences in economic geography that make some neighborhoods more likely to be treated than others.

[20] In principle, one could use either the construction year information or the first appearance in the panel for this purpose, but there are significant inconsistencies in both sources of information. In particular, new constructions usually appear in the IPTU data one or two years after their actual conclusion.

Table 1: Summary Statistics: New Commercial Buildings

	Median	Mean	Std. Deviation	Min	Max
Total Land Area (m$ˆ2$)	3,622	8,292.4	11,971.2	750	54,082
Occupied Land Area (m$ˆ2$)	2,157	3,979.9	3,823	573	17,778
Built-Area-Ratio	10.4	12.6	8.1	0.7	48.2
Establishments	13.4	24.4	34	1	175.6
Workers	810.3	1,534.3	2,001.9	511	11,963.2
% College	61	58.9	17	31.6	89.7
% of High-skilled Office workers	35.5	39.6	27	0	99.3
% of Low-skilled Office workers	5.3	14.6	21.8	0	94.6
% of Local Services workers	3.2	11.9	20.8	0	100
% of Non-office workers	26.1	33.7	29.7	0	99.8

Notes: The table displays summary statistics for the sample of new buildings used as local employment shocks. For time-varying characteristics such as employment and the number of establishments, quantities are computed based on average values observed after the inauguration.

Based on this new information, I establish the following criteria to select the "treatment" buildings:

- inauguration year (as defined below) between 2006 and 2013;
- at least 500 workers on average starting from the inauguration year;
- at least 25% of workers with a college degree on average starting from the inauguration year.

I define the inauguration year as the first period in which a building has 50 or more formal labor contracts in effect. This threshold is achieved in most cases in the first year after employment begins to kick in. Figure C2 shows that the occupation rate of these buildings accelerates rapidly after the inauguration.[21]

The second and third criteria impose that the shocks in employment are large enough so that local effects are potentially sizable. In particular, the threshold for individuals with a college degree speaks to the literature arguing that this group is more sensitive to agglomeration externalities (Moretti, 2004; Davis and Dingel, 2019).

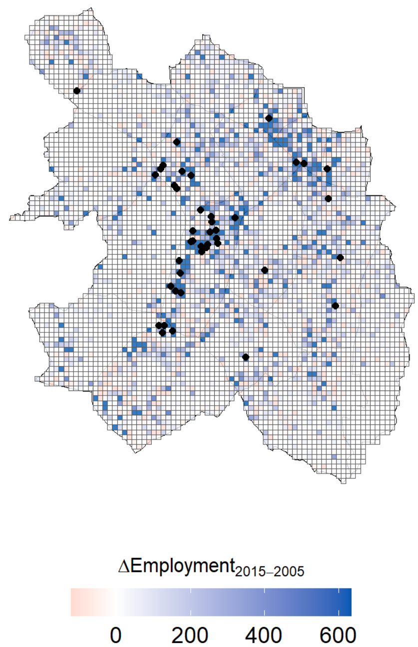
After applying these filters, I am left with 43 new commercial buildings, summarized in Table 1. Note that the average employment in these buildings ranges from a few more than 500 workers to almost 12,000. Regarding the number of establishments, there are buildings in which one big company settles, but in general, multiple establishments occupy the newly available space.[22]

Figure 3 shows the buildings' location and the variation in employment between 2015 and 2005. It is worth noting that employment has increased relatively more during this period in regions in the vicinity

[21]The matching between RAIS and IPTU by year reveals a few labor contracts associated with buildings before their inauguration. The establishments associated with these workers are usually local services or related to the construction sector.

[22]The timeline of openings is: 9 buildings in 2006, 4 in 2007, 5 in 2008, 3 in 2009, 8 in 2010, 3 in 2011, 4 in 2012 and 7 in 2013.

Figure 3: Location of New Commercial Buildings in Sao Paulo



Notes: The figure shows in black dots the location of the new commercial buildings considered in the analysis, together with cell variation in employment between 2005 and 2015.

of the buildings, which reinforces the possibility that their location correlates with trends in local employment. Together with Figure 1, it also shows that these buildings tend to be located close to high-density neighborhoods, suggesting that the relative spatial distribution of employment has not changed dramatically. Another aspect to note is that many neighborhoods observe multiple new buildings nearby. This fact creates a few challenges when building the econometric specification, which are addressed in the next section.

In order to validate the sample obtained, I use Google Maps to find the buildings and take screenshots of them. Some of these pictures are shown in Figure C3. Since the website makes available all imagery produced since 2010, it is possible for a subset of buildings (14 of 43) opened between 2011 and 2013 to check whether the inauguration year is consistent with what the images show over the years. Figure C4 illustrates one example: for a building whose inauguration year is 2013, as defined above, I observe that in 2011 the construction was still ongoing, but in 2014 it was already completed. For all buildings where this procedure can be carried out, the images are consistent with the inauguration year attributed to them.

4.2 Selecting Treated and Control Neighborhoods, Defining Treatment

After establishing the buildings to be used as local shocks, I define the sample of neighborhoods for the empirical analysis. Departing from the grid that covers Sao Paulo, I first select cells whose centroid is within

1 km of at least one new building's site. Then, I exclude cells that received any of the 43 new commercial buildings. To focus on changes at the intensive margin and guarantee that treatment effects across different sectors are comparable, I opt to limit the analysis to cells that contain at least one worker in each sector in all periods. Thus, the final sample is a balanced panel of 478 cells between 2003 and 2017.[23]

The next step is to define treatment and treated/control groups. A fundamental feature of my empirical setting, illustrated by Figure 3, is that the buildings are grouped in a few locations, so there are cells potentially affected by more than one building and in different magnitudes, depending on how close they are to new constructions. Hence, in order to build a standard staggered specification that can be appropriately estimated using up-to-date techniques, I create a classification of exposure to new buildings that defines two treated groups and one control:

- Treated Group 1 - first-ring cells: closest new building is within 0 to 250 meters (67 cells);
- Treated Group 2 - second-ring cells: closest new building is within 250 to 500 meters (135 cells);
- Control Group - outer-ring cells: closest new building is within 500 to 1000 meters (276 cells).

Moreover, I define the first treatment period as the year I observe the first building inauguration within the distance bin associated with the cell. For example, suppose a cell is close to three new commercial buildings, one inaugurated in 2008 that is 350 meters distant, one inaugurated in 2009 that is 140 meters distant and one inaugurated in 2012 that is 60 meters distant. In this case, the closest new building is within 0 to 250 meters; therefore, the cell is in Treated Group 1. The first treatment period is 2009 because it is the inauguration year of the first building within 0 and 250 meters. Figure 4 panel A shows the spatial distribution of these cells with the classification proposed.[24]

4.3 Predicting New Buildings' Locations

In order to make more accurate comparisons across treated and control groups, my research design also combines the "ring" approach with a procedure that explores the probability that a cell observes a new building in its vicinity. This probability comes from the estimation of the following propensity score model:

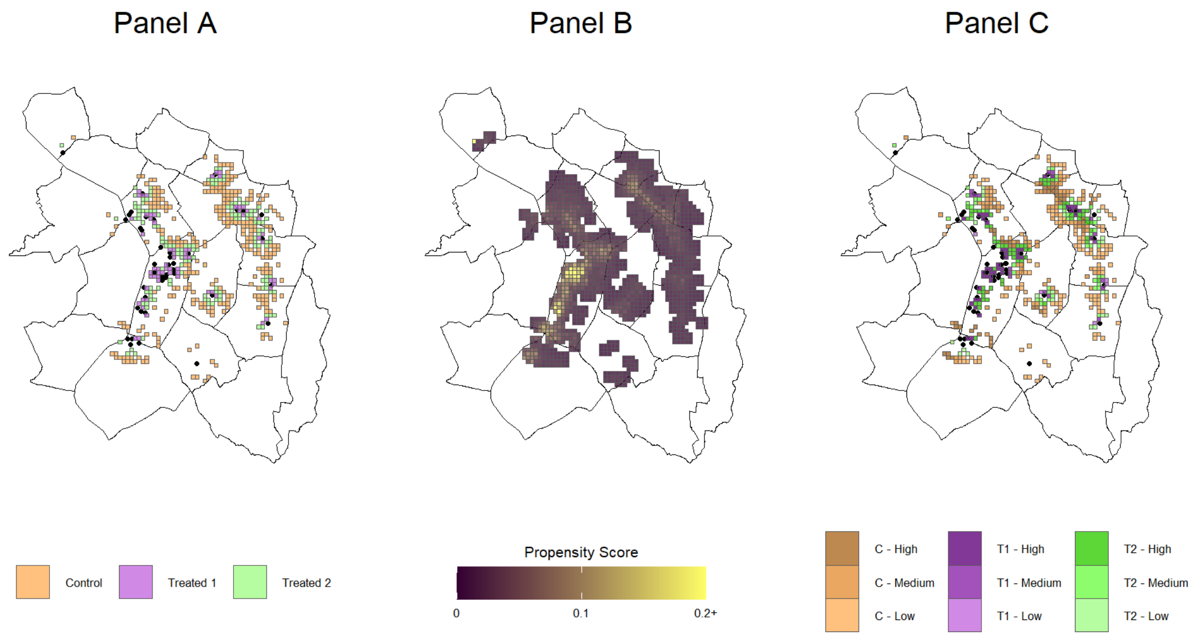
$$\mathbb{E}[Entry_c|X] = logistic(X_c\beta), \tag{12}$$

where $Entry_c$ is a binary variable that takes the value 1 if cell c receives at least one of the 43 new buildings considered in the analysis. X_c is a vector of variables that potentially predicts the construction

[23] As a robustness check, I provide results using alternative samples, reported in section ??.

[24] It is important to stress that the term "control group" does not mean that new buildings do not impact this group of cells. Strictly speaking, my empirical strategy aims to identify differential effects between nearer and farther neighborhoods.

Figure 4: Empirical Analysis Setup



Note: The figure depicts the design of the empirical analysis. Panel A shows the spatial distribution of treated and control cells. The black dots indicate the location of new commercial buildings. Panel B presents the results of the propensity score model estimated using Lasso. Finally, Panel C shows the treated/control classification together with the proximity-probability classification derived from the propensity score model. T1, T2 and C account for Treated Group 1, Treated Group 2 and Control Group, respectively, whereas High, Medium and Low refers to the PP terciles. Sections 4.2 and 4.3 provide the details.

of one of these buildings. It includes information on employment, wages, demography and transportation access from different sources. Some of these variables are included in level and variation prior to 2006. Table C2 shows the complete list.

Equation (12) is estimated using Lasso. For this purpose, I use an extended sample that contains treated/control cells described in the previous section, cells that received a new building and a group of" peripheral" cells located within 250 meters from a treated/control cell. The reason for including the last group will be explained below. Table C1 presents the parameters obtained for the chosen lambda, and Figure 4 panel B exhibits the spatial distribution of fitted values.[25]

After this procedure, I compute the average propensity score in a 250 m radius circumference drawn from the centroid of each treated/control cell. This is the reason for including peripheral cells in the estimation. I interpret the obtained values as measures related to the probability of being close to a new building, henceforth Proximity Probability (PP). Then, I divide cells in terciles based on their PP level. Panel C of Figure 4 plots the final result of this procedure, with the combined PP and treated/control classification.

In Table 2, I display summary statistics for the cell groups defined above using the main outcome variables

[25]The lambda parameter is determined using a 10-fold cross-validation.

analyzed in this study (all quantities represent simple averages). It is worth noting that first-ring cells with high PP are much more similar to second- and outer-ring cells within the same PP group than first-ring cells in other PP groups, thus showing that the propensity score method does a fairly good job of identifying more appropriate control groups.

4.4 Econometric Specifications

To evaluate the effects of new commercial buildings on an outcome y in cell c and year t, the equation to be estimated is

$$y_{c,t} = \sum_{k=-4}^{5} [\alpha_{k,250} D_{c,k,t,250} + \alpha_{k,500} D_{c,k,t,500}] + \Psi_c + \xi_{PP,t} + \mu_{D,t} + \epsilon_{c,t} \quad . \tag{13}$$

In this expression, the subscripts 250 and 500 allude to the 0-250m and 250-500m treated groups, respectively. The subscript k represents event periods relative to t.

The treatment variable $D_{c,k,t,d}$ takes the value 1 if cell c is in the treated group d and if the difference between year t and the year of treatment adoption is k. Thus, $\alpha_{k,d}$ represents the average effect of being differentially exposed to new buildings k periods from the start of treatment for group d. The coefficients related to $k = -1$ are normalized to zero.[26]

The specification also includes cell fixed-effects Ψ_c to control for invariant neighborhood characteristics. Finally, the coefficients $\xi_{PP,t}$ and $\mu_{D,t}$ are the interaction of PP level and time indicators and the interaction of district and time indicators, respectively. Thus, treatment effects are estimated controlling for district-level trends and using variation only within groups of cells with a similar likelihood to be close to a new commercial building, as informed by the propensity score model.

I also estimate average treatment effects using a standard static model. This specification significantly reduces the number of parameters of interest and provides better-powered estimates that are simpler to interpret:

$$y_{c,t} = \alpha_{250} D_{c,t,250} + \alpha_{500} D_{c,t,500} + \Psi_c + \xi_{PP,t} + \mu_{D,t} + \upsilon_{c,t} \quad , \tag{14}$$

where $D_{c,t,250}$ and $D_{c,t,500}$ are indicators of whether cell c received one of the treatments in period t, and α_{250} and α_{500} are the treatment effects now averaged in time.

I estimate Equations (13) and (14) using the two-stage procedure proposed by Gardner (2021), which consists of regressing $y_{c,t}$ on $\Psi_c + \xi_{PP,t} + \mu_{D,t}$ using only the untreated observations and then regressing

[26] If k is lower than -4 or greater than 5, I consider that $D_{c,-4,t,d} = 1$ and $D_{c,5,t,d} = 1$, respectively, i.e., those event periods are "binned".

Table 2: Summary Statistics: Cells

	T1 - High	T1 - Medium	T1 - Low	T2 - High	T2 - Medium	T2 - Low	C - High	C - Medium	C - Low
Workers									
High-Skilled Offices	715.9	90.6	122	520	113	54.3	569.3	124	144.9
Low-Skilled Offices	373.3	164	16.3	497	476.5	72	461.4	184.3	86.6
Local Services	441.6	172.6	249.3	238.9	201.5	190.1	467.9	145.6	117.8
Non-offices	567.8	200	70.1	401.1	161.1	107.3	359	141.8	113
Establishments									
High-Skilled Offices	33.7	9.6	4.3	23.4	11.4	4.6	24.3	8.9	6.1
Low-Skilled Offices	22.3	10	5.7	20	17.1	7.3	18.8	13.6	7.8
Local Services	38.2	19.9	18.3	27	23.3	18.7	41.6	20.9	15.7
Non-offices	22.1	9.3	4.6	17.4	11.5	7.5	17.3	9.3	7.3
Wages									
High-Skilled Offices	7,234	3,237.4	3,527.5	5,029.9	3,755.7	3,133.3	6,095.4	3,533.5	3,322.7
Low-Skilled Offices	4,499.6	2,001.8	1,616.2	3,400	2,627	2,222.2	3,438.7	2,455.1	2,269.6
Local Services	3,280.4	2,120.8	2,523.8	2,750.7	2,622	1,923.5	3,012	2,184.1	1,904.2
Non-offices	7,497	3,267.1	3,072.6	5,141.8	3,852.3	2,969.1	5,769.1	3,764.4	3,394.3
% College graduates									
High-Skilled Offices	52.2	26.9	31.9	42.5	33.9	28.5	46.9	33	30.3
Low-Skilled Offices	32.9	14.8	20.7	24.2	19.3	16.5	24.3	18	15.9
Local Services	17.2	10.7	13.6	13.6	14.4	8.2	15.5	10.3	8.3
Non-offices	38.8	21.1	19.6	30.7	25.2	22.8	34.2	25.8	23.2

Notes: The table displays summary statistics for different cell groups used in the empirical analysis. T1, T2 and C account for Treated Group 1, Treated Group 2 and Control Group, respectively, whereas High, Medium and Low refers to the PP terciles. Each value is an average for the quantity indicated in the first column in the left.

the adjusted outcomes $y_{c,t} - \hat{\Psi}_c + \hat{\xi}_{PP,t} + \hat{\mu}_{D,t}$ on $D_{c,t,250}$ and $D_{c,t,500}$ (or $D_{c,k,t,250}$ and $D_{c,k,t,500}$ when estimating the event-study). Under parallel trends and no anticipation assumptions, this approach delivers estimates robust to heterogeneous treatment effects over cells and periods. Standard errors are clustered at the cell level.[27,28]

5 Results

I now present estimates of new building effects for various outcomes and discuss how they relate to the theoretical predictions of Section 3. I start with event-study specifications (Equation 13) to analyze qualitative changes and check for pre-trends, and then move to the static specifications (Equation 14) to better understand the effects quantitatively. In the end, I discuss alternative interpretations of the results.

5.1 Event Study

I start by analyzing the effects of new commercial buildings on the aggregate level of economic activity. Figure 5 displays plots of results from Equation (13) for various outcomes. The upper left and right panels show, respectively, that the number of establishments and workers in first-ring cells are differentially impacted, and the positive effects persist over the period of analysis. Importantly, these panels show no sign of pre-trends prior to the beginning of treatment.

In the lower left and right panels of Figure 5, I display the effects of new buildings on the average wage premium and the share of college-educated workers, respectively, two measures potentially correlated with local productivity. They show no evidence that treated cells were affected.

While the upper panels indicate that the level of economic activity in first-ring cells is differentially affected, the lower panels suggest that local productivity remains stable in treated cells. According to the theoretical discussion in Section 3, we should expect both variables to be affected if the employment shocks associated with the new buildings generate spillovers on productivity. However, it is possible that such effects are concentrated on specific sectors, which may be indistinguishable when analyzing the aggregate economy.

Hence, I now investigate how each of the four sectors I defined has responded in treated cells. Figure 6 shows a differential expansion in the number of establishments providing local services that takes effect in the first period after the treatment's initiation. Despite not being possible to reject the null hypothesis at the 5% significance level, the results also suggest an increase in the number of high- and low-skilled offices.

[27] Borusyak et al. (2021), Liu et al. (2022), and Wooldridge (2021) propose similar estimators with minor differences between them. One feature that makes this method more suitable to my setting is that it allows me to account for specific trends that might confound the results. See de Chaisemartin and D'Haultfoeuille (2022) for a detailed discussion.

[28] Table D8 exhibits results using an alternative clustering based on the closest commercial building.

Figure 7 shows a more striking pattern in terms of employment, with strong and persistent effects for both local services and high-skilled offices.

The employment growth of these two sectors in first-ring cells appears to explain the bulk of the effects on the overall economy reported in Figure 5. Through the lens of the model, it can be interpreted as a manifestation of the two types of agglomeration forces considered in this paper. The higher presence in high-skilled offices would indicate that new buildings boost local productivity through spillovers, whereas the higher presence of local services would indicate an increase in local demand for non-tradable goods.

Next, I show how wages and the share of college-educated workers were affected by sector. Figure 8 suggests that the mean wage premium increases differentially in first-ring cells for high-skilled offices but not for other sectors. There is a rise in point estimates after the treatment starts that persists over the period of analysis, although it is not possible to reject the null hypothesis at the 5% significance level. High-skilled offices also experience an increase in the share of college-educated workers in first-ring cells, as displayed in Figure 9. In this case, the non-office sector seems to be impacted as well.

The effects on high-skilled offices are consistent with the predictions of the model. If new buildings trigger a shift in the local productivity of this sector, we should expect wages to increase. On the other hand, the model also predicts that the wages paid by local services would increase, but Figure 8 shows no evidence of that. One possible explanation for this fact is the existence of heterogeneities in labor supply not considered in the model. If different sectors employ different types of workers and local services observe a more elastic labor supply curve, this sector may experience a negligible impact on wages.

5.2 Standard Static Difference-in-Differences

I now examine the results from Equation (14) on the sector outcomes considered so far, which are summarized in Table 3. Panel A displays the effects on the aggregate economy. Consistent with the event-study plots, Columns (1) and (2) exhibit impacts of 4.3% and 11.1% in the number of establishments and workers, respectively. Likewise, Columns (3) and (5) show that the share of college-educated workers and wages were unaffected.

Regarding high-skilled offices, Panel B shows no evidence of effects in the number of establishments but an increase of 20.9% in employment in first-ring cells. This sector also experiences an increase of 3.4 percentage points in the share of college-educated workers and an increase of 6.1% in the average wage premium.

Panel D presents evidence that the presence of local services is impacted as well. There is a 7.7% increase in establishments and a 13.4% increase in employment in first-ring cells. On the other hand, there is no evidence that wages or the share of workers with a college degree are affected.

Table 3: Effects of New Commercial Buildings: Standard DiD

	Log Estabs	Log Workers	% College		Wage Premium	
			All estabs	excl. new estabs	All estabs	excl. new estabs
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. All Sectors						
0-250m	0.0430*	0.1111**	0.0078		0.0112	
	(0.0241)	(0.0484)	(0.0097)		(0.0147)	
250-500m	0.0058	0.0227	-0.0038		-0.0006	
	(0.0133)	(0.0347)	(0.0060)		(0.0100)	
R^2	0.00680	0.00658	0.00189		0.00080	
Observations	7,170	7,170	7,170		7,170	
Panel B. High-Skilled Offices						
0-250m	0.0495	0.2087**	0.0337**	0.0231*	0.0609**	0.0186
	(0.0417)	(0.0842)	(0.0135)	(0.0137)	(0.0251)	(0.0206)
250-500m	0.0260	0.0920	0.0043	-0.0016	0.0210	0.0088
	(0.0294)	(0.0659)	(0.0088)	(0.0099)	(0.0148)	(0.0152)
R^2	0.00206	0.00765	0.00620	0.00332	0.00873	0.00093
Observations	7,170	7,170	7,170	6,705	7,170	6,705
Panel C. Low-Skilled Offices						
0-250m	0.0677*	0.0743	0.0040	-0.0128	-0.0368*	-0.0464**
	(0.0372)	(0.1120)	(0.0129)	(0.0127)	(0.0197)	(0.0195)
250-500m	0.0443*	0.1035	-0.0073	-0.0174*	-0.0041	-0.0515***
	(0.0262)	(0.0724)	(0.0097)	(0.0098)	(0.0137)	(0.0132)
R^2	0.00563	0.00218	0.00074	0.00364	0.00367	0.01559
Observations	7,170	7,170	7,170	6,765	7,170	6,765
Panel D. Local Services						
0-250m	0.0770**	0.1341**	-0.0067	-0.0157**	-0.0107	-0.0335**
	(0.0374)	(0.0555)	(0.0067)	(0.0076)	(0.0112)	(0.0135)
250-500m	-0.0034	-0.0016	-0.0077	-0.0096*	-0.0127	-0.0183**
	(0.0186)	(0.0314)	(0.0056)	(0.0055)	(0.0086)	(0.0089)
R^2	0.01093	0.01076	0.00230	0.00640	0.00223	0.00909
Observations	7,170	7,170	7,170	7,080	7,170	7,080
Panel E. Non-Offices						
0-250m	0.0016	0.0798	0.0258**	0.0258*	0.0142	0.0030
	(0.0447)	(0.0781)	(0.0116)	(0.0134)	(0.0206)	(0.0191)
250-500m	-0.0109	-0.0292	-0.0032	0.0048	-0.0069	-0.0335**
	(0.0238)	(0.0474)	(0.0086)	(0.0100)	(0.0135)	(0.0153)
R^2	0.000083	0.00219	0.00579	0.00420	0.00095	0.00555
Observations	7,170	7,170	7,170	6,945	7,170	6,945

Notes: This table reports estimates of α_{250} and α_{500} in Equation (14) for different outcome variables indicated in the columns. Standard errors clustered at the cell level are displayed in parentheses. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

It is worthwhile to note that the expansion of local services and high-skilled offices in first-ring cells occurs in different ways. Local services increase their presence at the intensive and extensive margin, i.e., more establishments and more employment per establishment. However, for high-skilled offices, the growth happens mainly at the intensive margin, as the effect on the number of establishments for this sector does not seem to change meaningfully.

From the coefficients shown in Table 3, it is possible to derive a multiplier effect of the tradable sector on the non-tradable sector. Considering the baseline values, first-ring cells observe an increase of about 44 workers in local services and 97 workers in high-skilled offices, meaning that roughly one job is created in local services for every two additional jobs in high-skilled offices. This number is three times lower than the one obtained by Moretti (2010). However, this comparison should be made cautiously, as my definition of a non-tradable good is more strict.

There are two important caveats to consider in my assessment of the multiplier effect. First, it abstracts from the possibility that workers located in one neighborhood consume local services in the surrounding neighborhoods. In general, neglecting this issue would not alter the conclusion if these interconnections are symmetric, except when considering new buildings' direct effect on local demand. Given that new buildings are mostly occupied by offices, there is a large demand shock for local services that might be partially supplied by other neighborhoods, thus resulting in a potential overestimation of the multiplier effect (recall that neighborhoods that receive a new building are dropped from the estimation).

Secondly, Sao Paulo has a significant number of informal firms not included in the analysis. Given that informality is most likely more pervasive among local services, the employment effects on this sector (and the multiplier effect, as a consequence) may be underestimated.

5.3 Changes in Firm Composition Within Sectors

In Columns (4) and (6), I expand the analysis and investigate the effects on the share of college workers and the wage premium considering only firms located in treated neighborhoods before the treatment starts. Differences in the estimates between Columns (3) and (4) and between Columns (5) and (6) would inform whether the effects on these variables can be associated with changes in the composition of firms within sectors. Note that in these estimations, there is a small reduction in sample size since some neighborhoods have zero "old" firms in certain years and are dropped from the panel.

In both cases, there is a decrease in point estimates related to high-skilled offices in first-ring cells. Column (4) shows a decrease in treatment effect of 30% in comparison to Column (3), whereas Column (6) shows a decrease of 70% in comparison to Column (5). Moreover, the coefficient on the average wage premium is no longer significant at 10%.

This evidence suggests that changes in firm composition are relevant to understand why wages increase in high-skilled offices. Two effects might be at play, or a combination of both. One possibility is a higher inflow of highly productive firms that rely more heavily on college-educated workers and pay higher wages. The second possibility is a higher outflow of low-quality firms, perhaps due to increased rental costs.[29]

5.4 Floor Space Supply

As the number of establishments differentially increases in first-ring cells (driven mainly by local services), it is worth asking if this shift is followed by an increase in floor space supply. To check for this possibility, I estimate the effect of new buildings on the stock of commercial floor space computed from IPTU data. Figure 10 shows no evidence of an impact in the period analyzed.[30]

In the model, I allow the supply of floor space to adjust in response to changes in demand. However, this result indicates that the increase in establishments likely occurs through a decrease in vacancy rates. If this is indeed the case, a model of the floor space market with frictions may better capture the dynamics of this sector.

5.5 Productivity Spillovers

The results show that urban concentration is driven both by firms that produce non-tradable goods and therefore depend on local demand and by firms that produce tradable goods and whose demand is not concentrated in its vicinity. The increasing presence of the latter is less obvious and is consistent with the existence of local productivity spillovers. This interpretation becomes more plausible if we consider that the employment impact on high-skilled offices is accompanied by effects on wages and worker composition that do not occur in other sectors.

At the same time, the evidence also gives empirical support to rule out some alternative explanations. For instance, one could argue that the sector classification proposed in this paper is misleading, and local demand linkages between firms may instead explain the effects on high-skilled offices. However, it is difficult to reconcile this hypothesis with the fact that low-skilled offices are much less affected than high-skilled offices, as they would likely rely on similar linkages.

However, it is conceivable that being in a high-employment neighborhood raises the likelihood of closing

[29] Unfortunately, there is no reliable information on floor space value at a 200-meter grid cell level for the period I analyze. The IPTU data provides a measure of floor space value that reasonably correlates with market prices in a cross-section analysis. However, this information is mainly used for tax purposes and its evolution is highly influenced by political factors, thus doing a poor job in capturing price dynamics.

[30] Since, in this case, I have a panel that ends in 2019, I estimate event-study parameters up to $k = 7$.

more deals for high-skilled offices in other ways. For example, if physical proximity attenuates information frictions between firms and potential clients, or if being in an expensive location improves a firm's reputation, then location might be a crucial factor influencing demand in this sector. Nevertheless, while these mechanisms do not fit into the productivity spillover hypothesis, they are still based on externalities enhanced by density.

Another possible explanation is that neighborhoods closer to new buildings may become more appealing to workers. For example, if new buildings trigger improvements in local amenities or if their location is closely connected with improvements in labor market access (e.g., public investments in infrastructure). In this case, more firms would choose to locate in these neighborhoods as they would find it easier to recruit workers. High-skilled offices might be particularly benefited since they are more reliant on skilled workers, which are scarcer. However, in this scenario, wages are expected to decrease according to the model, a prediction that contrasts with the results.[31]

6 Robustness Checks

This section summarizes a series of tests to validate the robustness of the findings reported in this paper. Appendix D provides more details. Some specific exercises show lower employment effects on local services (which might suggest an overestimation of the multiplier effect) and lower wage effects on high-skilled offices, but overall the main results remain consistent.

6.1 Larger Sample of Neighborhoods

The first concern I address is related to the sample selection of neighborhoods. As explained in Section 4.2, I opt to restrict the analysis to neighborhoods with non-zero workers in all sectors and years to ensure comparable results. Since this choice might be overly restrictive, I report here additional results relaxing this requirement.

Departing again from the cells within 1 km of a new building site, I now select all cells with at least one worker in high- or low-skilled offices in all periods. I identify the treated and control groups similarly and re-estimate the propensity score using the new sample. Figure D1 illustrates this procedure, and Table D1 presents the results from Equation (14), which replicates the structure of Table 3. Note that, in this case, the samples used in each estimation differ because some cells may not contain all sectors.[32]

[31] Equations 8 and 10 show that an increase in B_n would lead to an increase in rent prices while not affecting the price of the non-tradable good, which in turn decreases the wage set by firms according to Equation 4.

[32] Note also that, in this case, the number of observations in Column (1) is higher because the analysis includes establishments with zero employment records.

Some differences between the two sets of results are worth mentioning. First, there is a strong impact on the number of establishments in the high-skilled office sector. Moreover, there is weak evidence that their wages are affected. Finally, Table D1 reports considerably smaller effects on local services (establishments and workers). Although the elasticities in Table D1 suggest that the multiplier effect is smaller than what is reported in Section 5, recall that the samples used in each estimation are different, which prevents an accurate evaluation of the relative growth across sectors.

6.2 Alternative Thresholds - New Buildings

Because the selection of new buildings is based on *ad hoc* choices, it is also worthwhile to check whether the results are sensitive to alternative samples. To this end, I perform the same analysis described in Section 4, now imposing different thresholds related to the average employment and the average share of college-degree workers.

Table D2 displays estimates of α_{250} from Equation (14) for different thresholds indicated in the columns. Column (1) replicates the baseline results when the average employment threshold n is 500 and the average share of college-degree workers sh is 25%. In Columns (2) to (5), I consider different values of n and sh. Overall, the results do not change substantially. In all cases, there is an increase in economic activity driven by local services and high-skilled offices. The latter also experience increases in wages and the share of college-degree workers.

6.3 Nearest Neighbor Matching using the Proximity Probability Measure

Dividing neighborhoods on terciles based on a propensity score can also raise doubts about the sensitiveness of the results to different settings. To address this issue, I report estimates using a standard nearest-neighbor matching with replacement based on the proximity probability measure developed. Departing from the sample described in Section 4.2, I pair each first- and second-ring cell with an outer-ring cell and estimate treatment effects on each group separately using Equation (14). Figure D2 shows the final samples obtained, and Tables D3 and D4 exhibit the results.

Compared to Table 3, the effect on the number of establishments in the local services (high-skilled office) sector is less (more) pronounced. The wage and worker composition effects are still concentrated on high-skilled offices. It is also worth noting that while this method offers more transparent comparisons among treated and control units, the estimations have less statistical power compared to the main results.

6.4 Other tests

Appendix D also provides results excluding second-ring cells from the estimation (Table D5), using continuous treatment based on the distance to new buildings (Table D6), using only cell and district-year fixed-effects as controls (Table D7) and an alternative clustering of standard errors based on the closest new building (Table D8).

Overall, the conclusions remain unchanged, with only a few exceptions. When considering continuous treatment effects, the impact on local services' employment appears to be smaller. Similarly, excluding PP-year fixed-effects from the regression results in a decrease in wage effects on high-skilled offices as well.

7 Conclusion

TBW

References

Ahlfeldt, G., Redding, S., Sturm, D., and Wolf, N. (2015). The economics of density: Evidence from the Berlin Wall. *Econometrica*, 83(6):2127–2189.

Arzaghi, M. and Henderson, J. (2008). Networking off Madison Avenue. *Review of Economic Studies*, 75:1011–1038.

Asquith, B., Mast, E., and Reed, D. (2021). Local effects of large new apartment buildings in low-income areas. *Working Paper*.

Baum-Snow, N., Gendron-Carrier, N., and Pavan, R. (2021). Local productivity spillovers. *Working Paper*.

Baum-Snow, N. and Pavan, R. (2012). Understanding the city size wage gap. *Review of Economic Studies*, 79(1):88–127.

Borusyak, K., Jaravel, X., and Spiess, J. (2021). Revisiting event study designs: Robust and efficient estimation. *arXiv preprint arXiv:2108–12419*.

Combes, P., Duranton, G., and Gobillon, L. (2008). Spatial wage disparities: Sorting matters. *Journal of Urban Economics*, 63:723–742.

Combes, P., Duranton, G., Gobillon, L., Puga, D., and Roux, S. (2012). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometria*, 80(6):2543—-2594.

Combes, P. and Gobillon, L. (2015). The empirics of agglomeration economies. In Duranton, G., Henderson, J., and Strange, W., editors, *Handbook of Regional and Urban Economics*, volume 5, pages 247–348. North Holland.

Davis, D. and Dingel, J. (2019). A spatial knowledge economy. *American Economic Review*, 109(1):153–170.

Davis, D. and Dingel, J. (2020). The comparative advantage of cities. *Journal of International Economics*, 123:103291.

Davis, D., Mengus, E., and Michalski, T. (2020). Labor market polarization and the great divergence: Theory and evidence. *NBER Working Paper no. 26955*.

de Chaisemartin, C. and D'Haultfoeuille, X. (2022). Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. *NBER Working Paper no. 29734*.

De La Roca, J. and Puga, D. (2017). Learning by working in big cities. *Review of Economic Studies*, 84(1):106–142.

Eckert, F., Ganapati, S., and Walsh, C. (2020). Skilled scalable services: The new urban bias in economic growth. *CESifo Working Paper No. 8705*.

Faber, B. and Gaubert, C. (2019). Tourism and economic development: Evidence from Mexico's coastline. *American Economic Review*, 109(6):2245—2293.

Gaubert, C. (2018). Firm sorting and agglomeration. *American Economic Review*, 108(11):3117–3153.

Greenstone, M., Hornbeck, R., and Moretti, E. (2010). Identifying agglomeration spillovers: Evidence from winners and losers of large plant. *Journal of Political Economy*, 118:536–598.

Heblich, S., Redding, S., and Sturm, D. (2020). The making of the modern metropolis: Evidence from London. *Quarterly Journal of Economics*, 135(4):2059—-2133.

Liu, C., Rosenthal, S., and Strange, W. (2020). Employment density and agglomeration economies in tall buildings. *Regional Science and Urban Economics*, 84:103555.

Liu, C., Rosenthal, S., and Strange, W. (2022). Agglomeration economies and the built environment: Evidence from specialized buildings and anchor tenants. *Working Paper*.

Moretti, E. (2004). Estimating the social return to higher education: Evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics*, 121:175–212.

Moretti, E. (2010). Local multipliers. *American Economic Review: Papers and Proceedings*, 100(2):373–377.

Moretti, E. (2011). Local labor markets. In Card, D. and Ashenfelter, O., editors, *Handbook of Labor Economics*, volume 4, pages 1237—-1313. North Holland.

Pennington, K. (2021). Does building new housing cause displacement? The supply and demand effects of construction in San Francisco. *SSRN Paper no. 3867764*.

Qian, F. and Tan, R. (2021). The effects of high-skilled firm entry on incumbent residents. *Working Paper*.

Redding, S. (2022). The economics of cities: From theory to data. *NBER Working Paper no. 30875*.

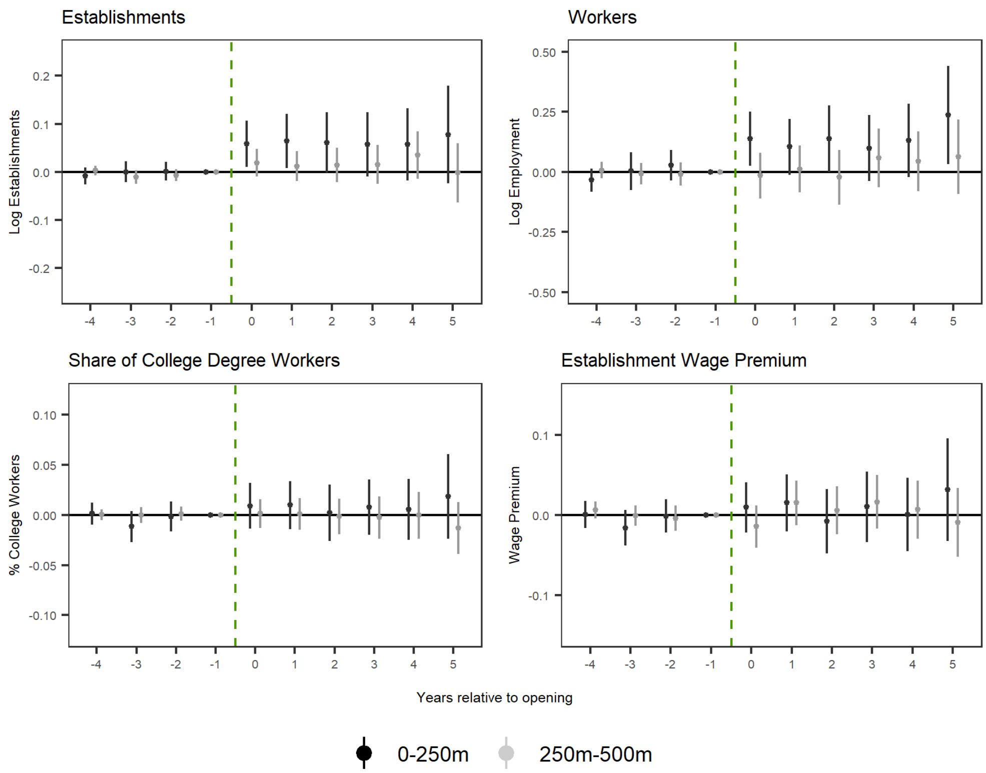
Rosenthal, S. and Strange, W. (2020). How close is close? The spatial reach of agglomeration economies. *Journal of Economic Perspectives*, 34(3):27–49.

Tsivanidis, N. (2022). Evaluating the impact of urban transit infrastructure: Evidence from Bogotá's Transmilenio. *Working Paper*.

Wooldridge, J. (2021). Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators. *SSRN Paper no. 3906345*.

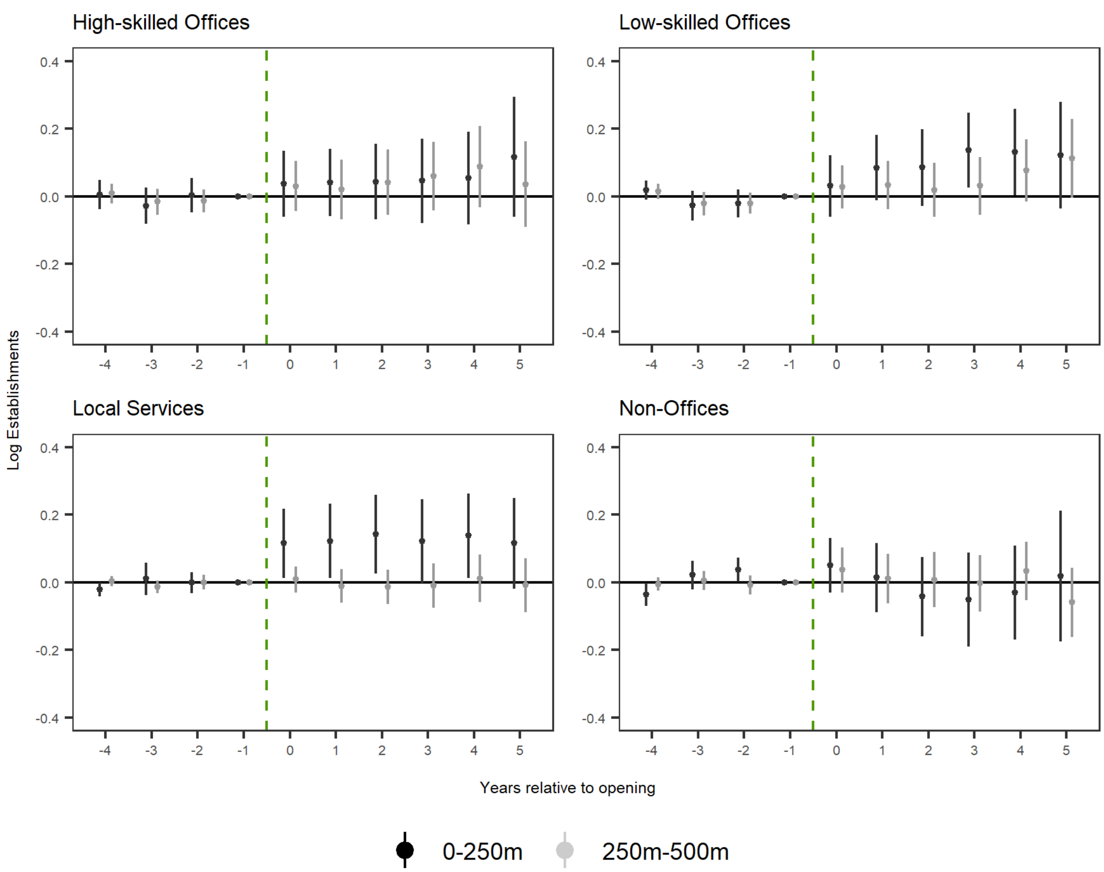
Figures and Tables

Figure 5: Event Study: Effects of New Buildings on Aggregate Outcomes



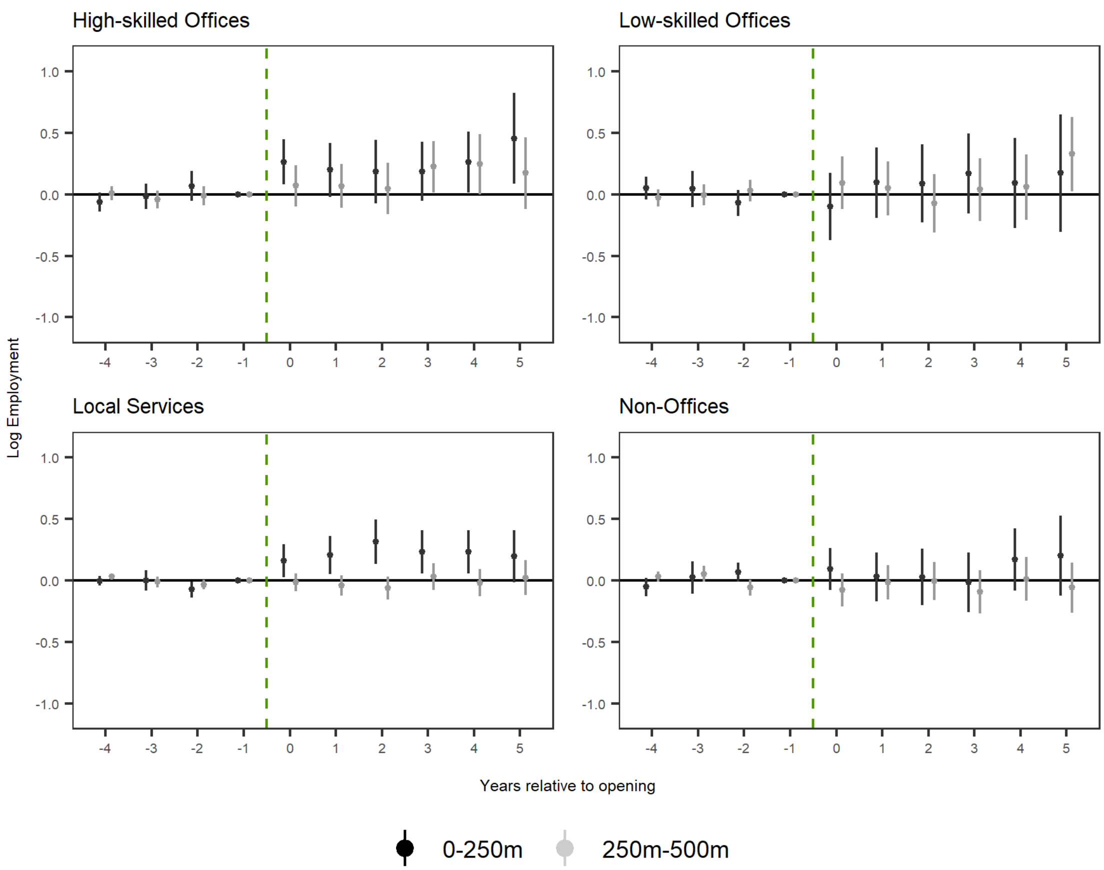
Note: figure plots coefficients from running Equation (13) on the log number of establishments (upper left panel), the log employment (upper right panel), the share of workers with college degree (lower left panel) and the wage premium (lower right panel). The bars indicates the 95% confidence interval, where standard errors are clustered at the cell level.

Figure 6: Event Study: Effects of New Buildings on the Number of Establishments By Sector



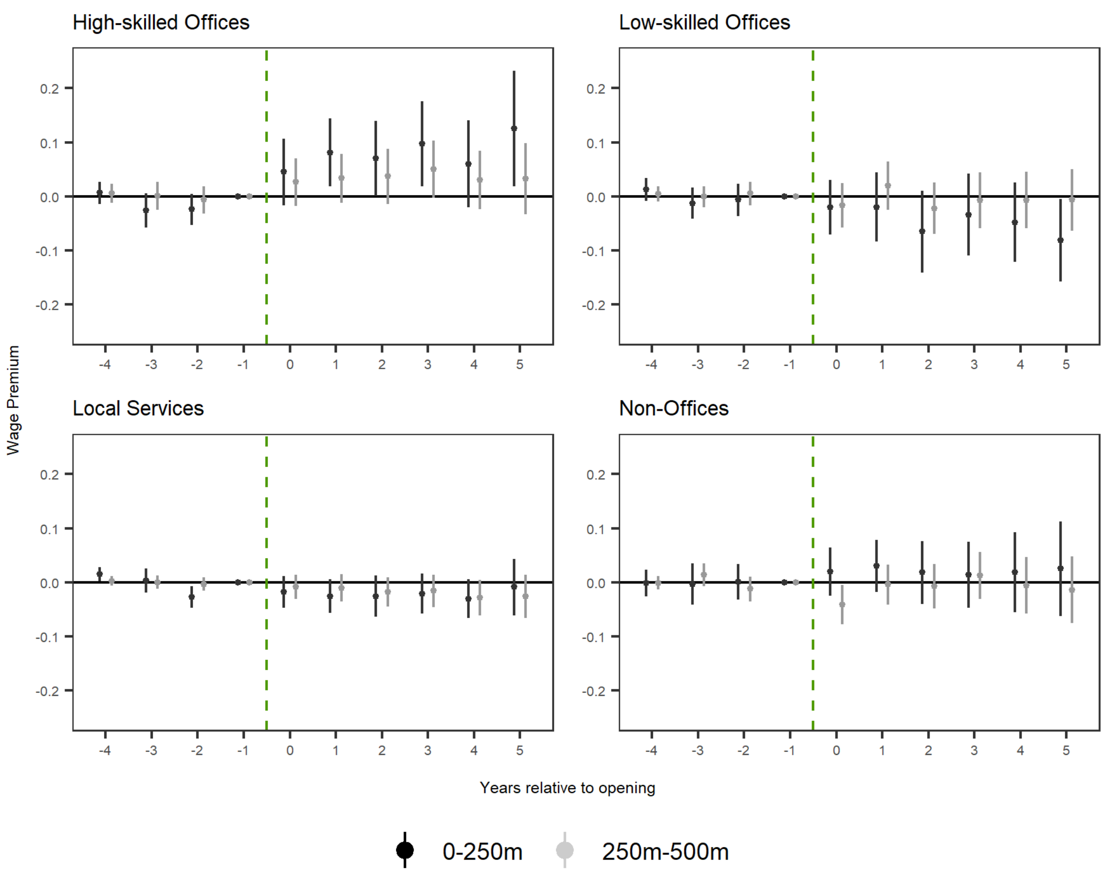
Note: figure plots coefficients from running Equation (13) on the log number of establishments for different sectors. The definition of each sector is described in Section 2.2. The bars indicates the 95% confidence interval, where standard errors are clustered at the cell level.

Figure 7: Event Study: Effects of New Buildings on Employment By Sector



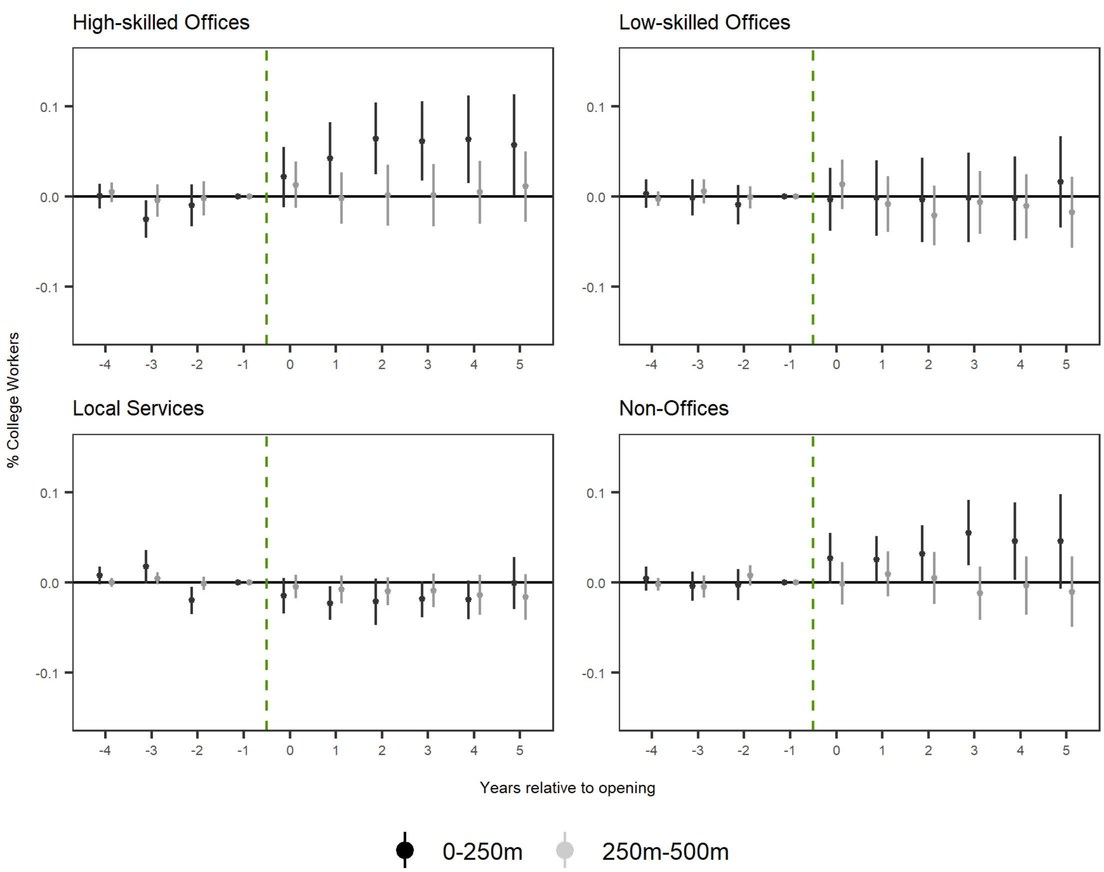
Note: figure plots coefficients from running Equation (13) on log employment for different sectors. The definition of each sector is described in Section 2.2. The bars indicates the 95% confidence interval, where standard errors are clustered at the cell level.

Figure 8: Event Study: Effects of New Buildings on the Average Wage Premium By Sector



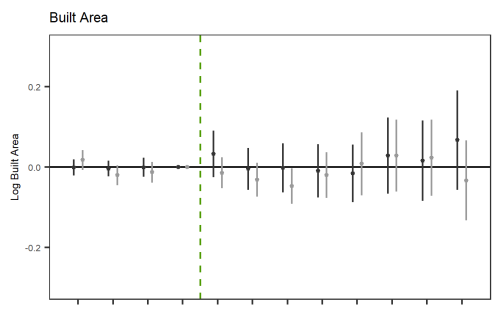
Note: figure plots coefficients from running Equation (13) on mean establishment wage premium (weighted by establishment size) for different sectors. The definition of each sector is described in Section 2.2. The bars indicates the 95% confidence interval, where standard errors are clustered at the cell level.

Figure 9: Event Study: Effects of New Buildings on the Share of College Workers



Note: figure plots coefficients from running Equation (13) on the share of workers with college degree for different sectors. The definition of each sector is described in Section 2.2. The bars indicates the 95% confidence interval, where standard errors are clustered at the cell level.

Figure 10: Event Study: Effects of New Buildings on the Supply of Commercial Floor Space



Note: figure plots coefficients from running Equation (13) on the log of commercial built area. The bars indicates the 95% confidence interval, where standard errors are clustered at the cell level.

Appendices

A - Multi-establishment Classification

For firms with at least 20 establishments located in Sao Paulo Metropolitan Area, I developed a procedure to obtain an accurate classification of establishments from the same firm. Consider a retail chain as an example. The goal is to distinguish a typical grocery store from administrative offices or distribution centers and categorize each type accordingly.

Using the establishment's identifier, it is possible to identify firms' headquarters and classify them as high-skilled offices. For non-HQ units, I perform an analysis based on their occupation composition. The intuition is that each firm can be characterized by one or two occupations with a significant presence in most establishments (e.g., cashiers in a retail chain).

I first use the 4-digit occupation code (CBO) to separate occupations between high- and low-skilled. The first group contains managers and professionals (CBO code < 3000). Then, for each establishment, I identify the occupation with the highest share and count how often each occupation is the most observed across establishments within the same firm. If an occupation is low-skilled and the top one in at least 10% of the establishments, I label it an essential occupation. For each firm, I select the two most important essential occupations. If only one occupation satisfies these conditions in a given firm, then only one occupation is selected.

The next step is to confront each establishment with the selected essential occupations. If they are above a threshold of 10% (i.e., if they are well represented), it means that the establishment is a typical one, so its classification is based on the 5-digit code (CNAE). However, if the share of main occupations is below the threshold, then the establishment is non-typical and needs another classification.

Next, I check if the non-typical establishments have at least 30 employees and 20% of high-skilled workers on average. If so, these establishments are likely administrative facilities, so I classify them as high-skilled offices. If one of these conditions is not satisfied, I classify the establishments as non-offices.

I validate this procedure using a sample of establishments from five firms: two commercial banks, two retail chains and a company that offers lab tests. Using address information, I search for the establishment on Google Street View and confront its facade with my classification.

B Theoretical Results and Derivations

B.1 Model Derivations

Firm Labor Supply.— Consider individuals i living in different neighborhoods m who need to choose a firm e to work. They take into account where the firm locates for two reasons. First, they spend a fraction δ of their wages on local services, whose price is neighborhood specific. Secondly, some neighborhoods offer higher utility than others.[33]

Let $n(e)$ be a function that maps the firm e with the neighborhood n where it is located. If the individual chooses to work in firm e, his indirect utility will be

$$u_{i,e} = B_{m,n(e)} \frac{w_e}{p^{\delta}_{LS,n(e)}} z_{i,e} \quad , \tag{B.1}$$

where $B_{m,n(e)}$ is how much individuals living in m value working in the neighborhood n where firm e is located and $z_{i,e}$ is an idiosyncratic shock of working in firm e. Individuals draw the idiosyncratic component independently for each firm from a Fréchet distribution whose cdf is $F^{ind}(z) = e^{-z^{-\eta}}$. As a consequence, the utility of an individual that lives in m working in e is also Fréchet distributed, and its cdf $G^{ind}_e(u)$ can be written as

$$G^{ind}_{m,e}(u) = F^{ind}\left(\frac{p_{LS,n(e)}}{B_{m,n(e)} w_e} u\right) = e^{-\phi_{m,e} u^{-\eta}} \quad ,$$

where $\phi_{m,e} = (B_{m,n(e)} w_e / p_{LS,n(e)})^{\tau}$. Using this distribution, it is possible to derive an expression for the probability $Pr_{m,e}$ that an individual from m chooses to work in firm e

$$Pr_{m,e} = \int_0^{\infty} Pr[u_e = max\{u_{e'}; \forall e'\}] dG^{ind}_{m,e}(u) \quad . \tag{B.2}$$

Now define $\phi_m \equiv \sum_{e'} \phi_{m,e'}$. This integral is solved by writing the term inside as a product of cdfs related to the utility distribution in all firms except e

[33] The reasoning behind this fact can be related either to amenities or to variations in commuting distance between neighborhoods.

$$
\begin{aligned}
Pr_{m,e} &= \int_0^\infty \prod_{e' \neq e} e^{-\phi_{m,e'}u^{-\eta}} (\phi_{m,e}\eta u^{-\eta-1}) e^{-\phi_{m,e}u^{-\eta}} du \\
&= \int_0^\infty (\phi_{m,e}\eta u^{-\eta-1}) e^{-\sum_{e'} \phi_{m,e'}u^{-\eta}} du = \\
&= \int_0^\infty \frac{d}{du}\left(\frac{\phi_{m,e}}{\phi_m} e^{-\phi_m u^{-\eta}}\right) du = \\
&= \frac{\phi_{m,e}}{\phi_m} = \frac{1}{\phi_m}\left(\frac{B_{m,n(e)}w_e}{p^{\delta}_{LS,n(e)}}\right)^{\eta} \quad .
\end{aligned}
$$

Assuming that each neighborhood has a fixed amount $\overline{L}_m$ of residents, the number of workers from m that choose to work in firm e is

$$
\ell_{m,e} = \frac{\overline{L}_m}{\phi_m}\left(\frac{B_{m,n(e)}w_e}{p^{\delta}_{LS,n(e)}}\right)^{\eta} \quad . \tag{B.3}
$$

Thus, the total number of workers that chooses to work in e can be computed by summing (B.3) over all neighborhoods m:

$$
\ell_e \equiv \sum_m \ell_{m,e} = \left(\frac{w_e}{p^{\delta}_{LS,n(e)}}\right)^{\eta} \sum_m \frac{\overline{L}_m}{\phi_m} B^{\eta}_{m,n(e)} \quad . \tag{B.4}
$$

Finally, I assume that the number of firms high enough so ϕ_m can be treated as fixed. I also $B_{n(e)} \equiv \sum_m \frac{\overline{L}_m}{\phi_m} B^{\tau}_{m,n(e)}$ the Firm Commuter Market Access of neighborhood n. Equation (3) is then obtained.

Firm Location Choice.— To derive Equations (5) and (6), it is necessary first to compute the respective profit functions. I do so by using the first-order conditions of (2), which yields

$$
\pi_{e,s,n} = \theta_{e,n}\left[\frac{1-\beta}{\beta}\frac{\varepsilon}{\varepsilon+1}\right]^{\varepsilon+1} \frac{\beta\chi}{\varepsilon} \cdot \begin{cases} \dfrac{B_n}{p^{\delta\varepsilon}_{LS,n}}\left(\dfrac{p_s A_{s,n}}{r^{\beta}_n}\right)^{\chi} & \text{if } s \neq LS \\ B_n \dfrac{p^{\chi-\delta\varepsilon}_{LS,n}}{r^{\chi\beta}_n} & \text{if } s = LS \end{cases} \quad . \tag{B.5}
$$

The rest of the derivation is analogous to the firm labor supply curve case. Profits for each pair sector-neighborhood follow a Fréchet distribution with cdf $G_{s,n}(\pi)$, which can be derived from $F(\theta)$. The expression that gives the probability that firm e chooses neighborhood n is obtained by solving an integral similar to (B.2). Finally, the number of firms from sector s that choose to locate in n is the product between this probability and $\overline{E}_s$

Floor Space Supply.— There is a construction sector that combines land T_n and capital K_n to produce floor space according to the following expression:

$$\max_{T_n, K_n} \quad \frac{r_n T_n^{\alpha} K_n^{1-\alpha}}{\alpha^{\alpha}(1-\alpha)^{(1-\alpha)}} - \rho_n T_n - K_n \quad \text{s.t.} \quad T_n \leq \overline{T}_n \quad , \tag{B.6}$$

where $\overline{T}_n$ is the total land available in n, and ρ_n is its price. The price of capital is uniform everywhere and assumed to be fixed and equal to 1. Using the first-order conditions, the local land supply curve in n is $\frac{\overline{T}_n}{\alpha} r_n^{\frac{1-\alpha}{\alpha}}$.

Market Clearing Conditions.— From the first order conditions of (2), it is possible to derive expressions for $w_{s,n}\ell_{s,n}$, $w_{LS,n}\ell_{LS,n}$ $Y_{LS,n}$, $f_{s,n}$ and $f_{LS,n}$:

$$w_{s,n}\ell_{s,n} = \left[\frac{\varepsilon}{\chi\beta}\right]^{\varepsilon+1} \beta^{\chi} \frac{B_n}{p_{LS,n}^{\delta\varepsilon}} \left(\frac{p_s A_{s,n}}{r_n^{\beta}}\right)^{\chi} \quad , \tag{B.7}$$

$$w_{LS,n}\ell_{LS,n} = \left[\frac{\varepsilon}{\chi\beta}\right]^{\varepsilon+1} \beta^{\chi} \frac{B_n p_{LS,n}^{\chi-\delta\varepsilon}}{r_n^{\chi\beta}} \quad , \tag{B.8}$$

$$Y_{LS,n} = \frac{w_{LS,n}\ell_{LS,n}}{p_{LS,n}} \frac{\chi}{\varepsilon} \quad \text{and} \tag{B.9}$$

$$f_{s,n} = \left[\frac{\varepsilon}{\chi\beta}\right]^{\varepsilon} \beta^{\chi} \frac{B_n}{r_n^{\chi-\varepsilon} p_{LS,n}^{\delta\varepsilon}} (p_s A_{s,n})^{\chi} \quad . \tag{B.10}$$

$$f_{LS,n} = \left[\frac{\varepsilon}{\chi\beta}\right]^{\varepsilon} \beta^{\chi} \frac{B_n p_{LS,n}^{\chi-\delta\varepsilon}}{r_n^{\chi-\varepsilon}} \quad . \tag{B.11}$$

Combining (B.9) and (7) yields

$$w_{LS,n}\ell_{LS,n}E_{LS,n} \left[\frac{\chi}{\delta\varepsilon} - 1\right] = \sum_{s \neq LS} w_{s,n}\ell_{s,n}E_{s,n},$$

Now, plugging (5), (6), (B.7) and (B.8) and solving for $p_{LS,n}$

$$p_{LS,n}^{\chi} p_{LS,n}^{\chi\eta} \frac{\overline{E}_{LS}}{\Phi_{LS}} \left[\frac{\chi}{\delta\varepsilon} - 1\right] = \sum_{s \neq LS} (p_s A_{s,n})^{\chi} A_{s,n}^{\chi\eta} \frac{\overline{E}_s}{\Phi_s}$$

$$\Longrightarrow p_{LS,n} = \left[\frac{1}{\chi/\delta\varepsilon - 1} \frac{\Phi_{LS}}{\overline{E}_{LS}} \sum_{s \neq LS} \frac{\overline{E}_s p_s^{\chi}}{\Phi_s} A_{s,n}^{\chi(1+\eta)} \right]^{\frac{1}{\chi(1+\eta)}}$$

To derive Equation (10), I first plug expressions (5), (6), (B.10) and (B.11) into (9):

$$\frac{\overline{T}_n}{\alpha} r_n^{\frac{1-\alpha}{\alpha}} = \left(\frac{\varepsilon}{\chi\beta}\right)^{\varepsilon} \beta^{\chi} \frac{B_n^{(1+\eta)}}{r_n^{\chi(1+\beta\eta)-\varepsilon} p_{LS,n}^{\delta\varepsilon(1+\eta)}} \left[\sum_{s \neq LS} \frac{\overline{E}_s p_s^{\chi}}{\Phi_s} A_{s,n}^{\chi(1+\eta)} + \frac{\overline{E}_{LS}}{\Phi_{LS}} p_{LS,n}^{\chi(1+\eta)} \right]$$

Now, note that Equation (8) can be used to substitute the summation term inside the brackets:

$$\frac{\overline{T}_n}{\alpha} r_n^{\frac{1-\alpha}{\alpha}} = \left(\frac{\varepsilon}{\chi\beta}\right)^{\varepsilon} \beta^{\chi} \frac{B_n^{(1+\eta)}}{r_n^{\chi(1+\beta\eta)-\varepsilon} p_{LS,n}^{\delta\varepsilon(1+\eta)}} \left[p_{LS,n}^{\chi(1+\eta)} \frac{\overline{E}_{LS}}{\Phi_{LS}} \left[\frac{\chi}{\delta\varepsilon} - 1\right] + \frac{\overline{E}_{LS}}{\Phi_{LS}} p_{LS,n}^{\chi(1+\eta)} \right]$$

$$\Longrightarrow \frac{\overline{T}_n}{\alpha} r_n^{\frac{1-\alpha}{\alpha}} = \frac{\chi}{\delta\varepsilon} \left(\frac{\varepsilon}{\chi\beta}\right)^{\varepsilon} \beta^{\chi} \frac{\overline{E}_{LS}}{\Phi_{LS}} \frac{B_n^{(1+\eta)} p_{LS,n}^{(\chi-\delta\varepsilon)(1+\eta)}}{r_n^{\chi(1+\beta\eta)-\varepsilon}}$$

Finally, I rearrange the terms to get

$$r_n = \left[\frac{\alpha}{\delta} \left(\frac{\chi}{\varepsilon}\right)^{1-\varepsilon} \beta^{\chi-\varepsilon} \frac{\overline{E}_{LS}}{\Phi_{LS}} \cdot \frac{B_n^{1+\eta}}{\overline{T}_n} \cdot p_{LS,n}^{(\chi-\delta\varepsilon)(1+\eta)} \right]^{\frac{1}{1/\alpha+\beta\chi(1+\eta)}} .$$

B.2 Proof of Proposition 1

To simplify notations, I first rewrite expressions (8) and (10) as:

$$p_{LS,n} = Z \left[\sum_{s \neq LS} \nu_s A_{s,n}^{\chi(1+\eta)} \right]^{\frac{1}{\chi(1+\eta)}} \tag{B.12}$$

and

$$r_n = Q_n p_{LS,n}^{\kappa} \quad , \tag{B.13}$$

where $\kappa = \dfrac{(\chi - \delta\varepsilon)(1+\eta)}{1/\alpha + \beta\chi(1+\eta)} > 0$. From these expressions, it is straightforward to show that $\dfrac{\partial p_{LS,n}}{\partial A_{s,n}}$ and $\dfrac{\partial r_n}{\partial A_{s,n}}$ are positive. Another useful expression to derive is the elasticity ξ of local services price with respect to $A_{s,n}$:

$$\xi \equiv \frac{A_{s,n}}{p_{LS,n}} \frac{\partial p_{LS,n}}{\partial A_{s,n}} = \frac{\nu_s A_{s,n}^{\chi(1+\eta)}}{\sum\limits_{s' \neq LS} \nu_{s'} A_{s',n}^{\chi(1+\eta)}} \quad , \tag{B.14}$$

Using the approximation $\dfrac{\partial \Phi_{s'}}{\partial A_{sn}} \approx 0 \ \forall \ s'$, proving that $\dfrac{\partial E_{s',n}}{\partial A_{s,n}} < 0$ for $s' \neq s$ is trivial from direct inspection of (5). Now, combining expressions (5) and (B.13) and taking the derivative of $E_{s,n}$ with respect to $A_{s,n}$:

$$\begin{aligned}
\frac{\partial E_{s,n}}{\partial A_{s,n}} &= \frac{\overline{E}_s}{\Phi_s} \left(\frac{B_n}{Q_n^{\chi\beta}}\right)^{\eta} \frac{\partial}{\partial A_{s,n}} \left(\frac{A_{s,n}}{p_{LS,n}^{\beta\kappa + \delta\varepsilon/\chi}}\right)^{\chi\eta} = \\
&= \frac{\overline{E}_s}{\Phi_s} \left(\frac{B_n}{Q_n^{\chi\beta}}\right)^{\eta} \chi\eta \left(\frac{A_{s,n}}{p_{LS,n}^{\beta\kappa + \delta\varepsilon/\chi}}\right)^{\chi\eta - 1} \left[\frac{1}{p_{LS,n}^{\beta\kappa + \delta\varepsilon/\chi}} - \left(\beta\kappa + \frac{\delta\varepsilon}{\chi}\right) \frac{A_{s,n}}{p_{LS,n}^{\beta\kappa + \delta\varepsilon/\chi - 1}} \frac{\partial p_{LS,n}}{\partial A_{s,n}}\right] = \\
&= \frac{\overline{E}_s}{\Phi_s} \left(\frac{B_n}{Q_n^{\chi\beta}}\right)^{\eta} \chi\eta \left(\frac{A_{s,n}}{p_{LS,n}^{\beta\kappa + \delta\varepsilon/\chi}}\right)^{\chi\eta - 1} \frac{1}{p_{LS,n}^{\beta\kappa + \delta\varepsilon/\chi}} \left[1 - \xi \left(\beta\kappa + \frac{\delta\varepsilon}{\chi}\right)\right]
\end{aligned}$$

Since $0 < \xi < 1$, the derivative is positive if $\beta\kappa + \dfrac{\delta\varepsilon}{\chi} < 1$. Using the definition of κ:

$$\begin{aligned}
\beta\kappa + \frac{\delta\varepsilon}{\chi} &= \beta \frac{(\chi - \delta\varepsilon)(1+\eta)}{1/\alpha + \beta\chi(1+\eta)} + \frac{\delta\varepsilon}{\chi} = \\
&= \frac{1 - \frac{\delta\varepsilon}{\chi}}{\frac{1}{\alpha\beta\chi(1+\eta)} + 1} + \frac{\delta\varepsilon}{\chi} < \\
&< \frac{1 - \frac{\delta\varepsilon}{\chi}}{1} + \frac{\delta\varepsilon}{\chi} = 1 \quad ,
\end{aligned}$$

and therefore $\dfrac{\partial E_{s,n}}{\partial A_{s,n}} > 0$.

For local services, I first combine Equations (6) and (B.13). Then, I take the derivative with respect to $A_{s,n}$ to get

$$\frac{\partial E_{LS,n}}{\partial A_{s,n}} = \frac{\overline{E}_{LS}}{\Phi_{LS}} \left(\frac{B_n}{Q_n^{\chi\beta}}\right)^{\eta} \frac{\partial}{\partial A_{s,n}} \left(p_{LS,n}^{1-(\beta\kappa + \delta\varepsilon/\chi)}\right)^{\chi\eta} \quad ,$$

and because $\beta\kappa + \dfrac{\delta\varepsilon}{\chi} < 1$, the derivative is positive.

B.3 Proof of Proposition 2

To simplify the notation, I rewrite (4) for an office sector as

$$w_{s,n} = M p_s^{\frac{1}{1-\beta}} \left(\frac{A_{s.n}}{r_n^\beta} \right)^{\frac{1}{1-\beta}} \quad , \tag{B.15}$$

Again using the approximation $\frac{\partial \Phi_{s'}}{\partial A_{sn}} \approx 0 \ \forall \ s'$, proving that $\frac{\partial E_{s',n}}{\partial A_{s,n}} < 0$ for $s' \neq s$ is trivial from direct inspection of (B.15). For sector s, the derivative of $w_{s,n}$ with respect to $A_{s,n}$ is

$$\begin{aligned}
\frac{\partial w_{s,n}}{\partial A_{s,n}} &= \frac{M}{1-\beta} p_s^{\frac{1}{1-\beta}} \left(\frac{A_{s.n}}{r_n^\beta} \right)^{\frac{1}{1-\beta}-1} \left(\frac{1}{r_n^\beta} - \beta \frac{A_{s.n}}{r_n^{\beta+1}} \frac{\partial r_n}{\partial A_{s,n}} \right) \\
&= \frac{M}{1-\beta} p_s^{\frac{1}{1-\beta}} \left(\frac{A_{s.n}}{r_n^\beta} \right)^{\frac{1}{1-\beta}-1} \frac{1}{r_n^\beta} \left(1 - \beta \frac{A_{s.n}}{r_n} \frac{\partial r_n}{\partial A_{s,n}} \right) \\
&= \frac{M}{1-\beta} p_s^{\frac{1}{1-\beta}} \left(\frac{A_{s.n}}{r_n^\beta} \right)^{\frac{1}{1-\beta}-1} \frac{1}{r_n^\beta} \left(1 - \beta\kappa\xi \right) \quad ,
\end{aligned}$$

where in the last row I use the fact that $\frac{\partial r_n}{\partial A_{s,n}} = \kappa\xi$, which can be easily proved from Equation (B.14). Since $0 < \xi < 1$ and $0 < \beta\kappa < 1$, this derivative is positive.

For local services, the same procedure yields

$$w_{LS,n} = M \left(p_{LS,n}^{1-\beta\kappa} \right)^{\frac{1}{1-\beta}} \quad , \tag{B.16}$$

and because $0 < \beta\kappa < 1$, $\frac{\partial w_{LS,n}}{\partial A_{s,n}}$ is positive.

For real wages, I first use Equations (B.15) and (B.16) to get expressions for $w_{s,n}/p_{LS,n}^\delta$ and $w_{LS,n}/p_{LS,n}^\delta$:

$$w_{s,n}^R \equiv \frac{w_{s,n}}{p_{LS,n}^\delta} = M p_s^{\frac{1}{1-\beta}} \left(\frac{A_{s.n}}{p_{LS,n}^{\beta\kappa+\delta(1-\beta)}} \right)^{\frac{1}{1-\beta}} \quad , \tag{B.17}$$

and

$$w_{LS,n}^R \equiv \frac{w_{LS,n}}{p_{LS,n}^\delta} = M \left(p_{LS,n}^{1-\beta\kappa-\delta(1-\beta)} \right)^{\frac{1}{1-\beta}} \quad , \tag{B.18}$$

Now, taking the derivative of $w_{s,n}^R$ with respect to $A_{s,n}$:

$$\frac{\partial w^R_{s,n}}{\partial A_{s,n}} = \frac{M}{1-\beta} p_s^{\frac{1}{1-\beta}} \left(\frac{A_{s.n}}{p_{LS,n}^{\beta\kappa+\delta(1-\beta)}} \right)^{\frac{1}{1-\beta}-1} \left[\frac{1}{p_{LS,n}^{\beta\kappa+\delta(1-\beta)}} - (\beta\kappa + \delta(1-\beta)) \frac{A_{s.n}}{p_{LS,n}^{\beta\kappa+\delta(1-\beta)+1}} \frac{\partial p_{LS,n}}{\partial A_{s,n}} \right]$$

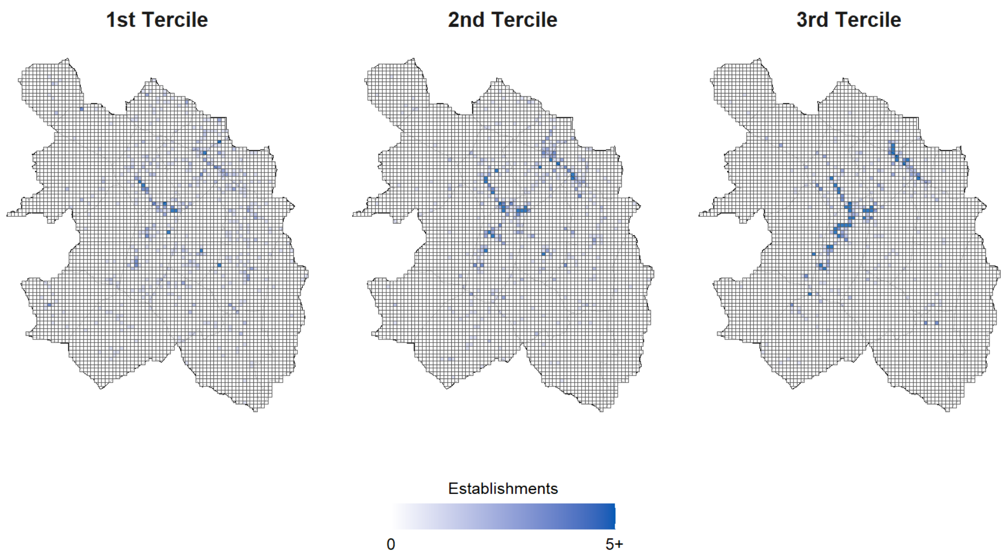
$$= \frac{M}{1-\beta} p_s^{\frac{1}{1-\beta}} \left(\frac{A_{s.n}}{p_{LS,n}^{\beta\kappa+\delta(1-\beta)}} \right)^{\frac{1}{1-\beta}-1} \frac{1}{p_{LS,n}^{\beta\kappa+\delta(1-\beta)}} \left[1 - (\beta\kappa + \delta(1-\beta))\xi \right] \quad ,$$

and the derivative is positive if $(\beta\kappa + \delta(1-\beta))\xi < 1$, which is not necessarily true.

For local services, direct inspection of Equation (B.18) shows that the derivative is positive if $\beta\kappa + \delta(1-\beta) < 1$, which is not guaranteed as well.

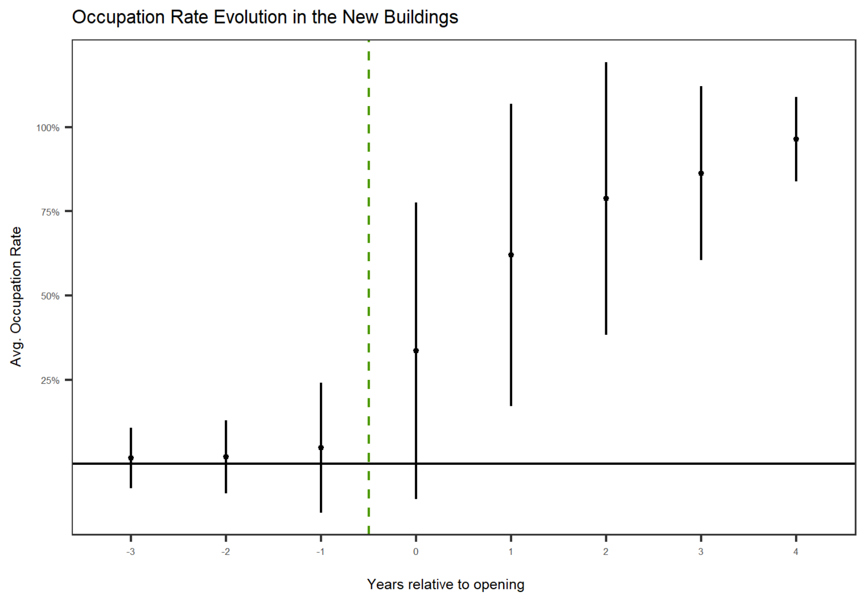
C Additional Figures and Tables

Figure C1: Spatial Distribution of Finance Establishments



Notes: The figure displays, for 2010, the spatial distribution of establishments in the financial industry by cell for different terciles of establishment wage premium. For more details about the estimation of these premia, see Section 2.1.

Figure C2: Evolution of Occupation Rate - New Buildings



Notes: The figure shows the average evolution of occupation rates after a new building is inaugurated. I define building capacity as the maximum number of workers observed. The bars indicates the 95% confidence interval.

Figure C3: Examples of Commercial Buildings

Source: Google Maps.

Figure C4: Validating Selected Buildings

Source: Google Maps. *Notes:* Example of a commercial building inaugurated in 2013. Using Google Maps imagery from 2011 and 2014, it is possible to check if the timeline of construction is consistent with the date of inauguration defined in the empirical strategy.

Table C1: Estimates of the Propensity Score Model

Variable	Estimate
Log built area growth commercial – cell	0.360
Log (1+ employment adm. and support) – cell	0.007
Log employment growth finance – cell	0.055
Log (1+ employment finance) – cell	0.025
Log (1+ employment information and communication) – buffer	0.098
Log (1+ employment information and communication) – cell	0.082
Log (1+ employment local services) – cell	0.006
Employment to population ratio	3.585
Log (1+ employment wholesale) – buffer	0.066
Log (1+ employment wholesale) – cell	0.072
% population 18-40 – buffer	1.132

Notes: Table shows the estimated coefficients of Equation (12) using Lasso. For a description of all variables used, see Table C2

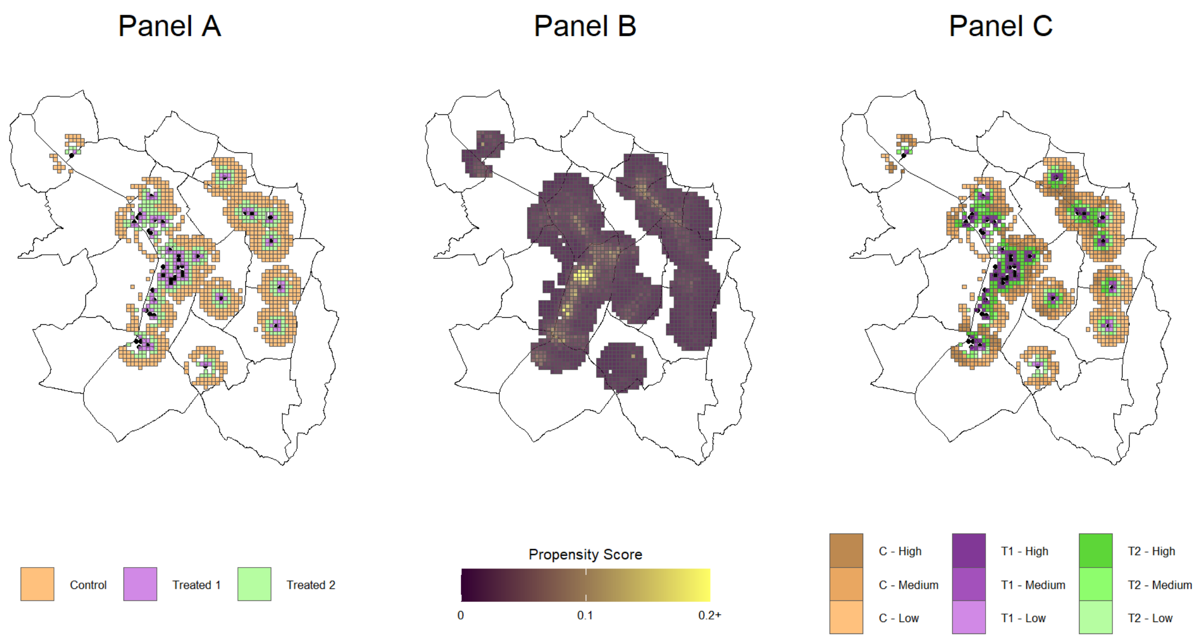
Table C2: List of Variables used in the Propensity Score Model

Variable Name	**Variable Description**	**Data Source**
Log (1 + employment) - cell	One plus the natural log of the number of formal workers in the cell.	RAIS 2005
Log (1 + employment) - buffer	One plus the natural log of the number of formal workers within 500 m from cell's centroid, excluding the cell itself.	RAIS 2005
Log (1 + employment industry) - cell	One plus the natural log of the number of formal workers by industry in the cell. Categories included: agriculture and manufacturing; transportation and utilities; professional services and real estate; construction; local services; wholesale; information and communication; finance; administrative and support; health; others categories.	RAIS 2005
Log (1 + employment industry) - buffer	One plus the natural log of the number of formal workers by sector within 500 m from cell's centroid, excluding the cell itself. For a list of industries, see third row.	RAIS 2005
Log mean wage - buffer	Natural log of the mean wage of formal workers located within 500 m from cell's centroid.	RAIS 2005
Log employment growth - cell	Difference in log of one plus employment between 2005 and 2003 in the cell.	RAIS 2003-2005
Log employment growth - buffer	Difference in log of one plus employment between 2005 and 2003 within 500 m from cell's centroid, excluding the cell itself.	RAIS 2003-2005
Log employment growth industry - cell	Difference in log of one plus employment by industry between 2005 and 2003 in the cell. For a list of industries, see third row.	RAIS 2003-2005
Log employment growth industry - buffer	Difference in log of one plus employment by industry between 2005 and 2003 within 500 m from cell's centroid, excluding the cell itself. For a list of industries, see third row.	RAIS 2003-2005
Log wage growth - buffer	Difference in log of the mean wage of formal workers between 2005 and 2003 within 500 m from cell's centroid.	RAIS 2003-2005
Log commercial built area - cell	Natural log of commercial stock of floor space in the cell	IPTU 2005
Log commercial built area growth - cell	Difference in log of commercial stock of floor space between 2005 and 2003 in the cell	IPTU 2003-2005
Log noncommercial built area - cell	Natural log of noncommercial stock of floor space in the cell	IPTU 2005
Log noncommercial built area growth - cell	Difference in log of noncommercial stock of floor space between 2005 and 2003 in the cell	IPTU 2003-2005
Number of train and subway stations - buffer	Number of train and subway stations within 500 m from cell's centroid.	SP Metro and CPTM
Log population - buffer	Natural log of number of residents within 500 m from cell's centroid.	2000 Census - tract level
Log households - buffer	Log number of households within 500 m from cell's centroid.	2000 Census - tract level
Log per capita income - buffer	Log per capita income within 500 m from cell's centroid.	2000 Census - tract level
% population 18-40 - buffer	Share of population between 18 and 40 years old within 500 m from cell's centroid.	2000 Census - tract level
% population 41-60 - buffer	Share of population between 41 and 60 years old within 500 m from cell's centroid.	2000 Census - tract level
% population non-white - buffer	Share of brown and black population within 500 m from cell's centroid.	2000 Census - tract level
% renters - buffer	Share of households that are renters within 500 m from cell's centroid.	2000 Census - tract level
% per capita income $<$ 1/4 of min. wage - buffer	Share of households whose per capita income is less than one quarter of a monthly minimum wage on within 500 m from cell's centroid.	2000 Census - tract level
% per capita income $>$ 1/4 and $<$ 1 min. wage - buffer	Share of households whose per capita income is greater than one quarter and less than one monthly minimum wage within 500 m from cell's centroid.	2000 Census - tract level
% per capita income $>$ 1 and $<$ 3 min. wages - buffer	Share of households whose per capita income is greater than one and less than three monthly minimum wages on census tracts within 500 m from cell's centroid.	2000 Census - tract level
% per capita income $>$ 3 min. wages - buffer	Share of households whose per capita income is greater than one and less than three monthly minimum wages within 500 m from cell's centroid.	2000 Census - tract level
Log distance to city center	Natural log of distance of cell's centroid to Se Square (in km).	-
Employment to population ratio - buffer	Ratio between employment and resident population within 500 m from cell's centroid.	RAIS 2005 and Census 2000 - tract level

D Robustness Checks

D.1 Larger Sample of Neighborhoods

Figure D1: Empirical Analysis Setup



Notes: the figure depicts the design of the empirical analysis using an alternative sample of cells, as described in Section 6.1. Panel A shows the spatial distribution of treated and control cells. The black dots indicate the location of new commercial buildings. Panel B presents the results of the propensity score model estimated using Lasso. Finally, Panel C shows the treated/control classification together with the proximity-probability classification derived from the propensity score model. T1, T2 and C account for Treated Group 1, Treated Group 2 and Control Group, respectively, whereas High, Medium and Low refers to the PP terciles. Sections 4.2 and 4.3 provide the details.

Table D1: Effects of New Commercial Buildings: Larger Sample of Neighborhoods

	Log Estabs	Log Workers	% College		Wage Premium	
			All estabs	excl. new estabs	All estabs	excl. new estabs
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. All Sectors						
0-250m	0.0572***	0.0925**	0.0109		0.0086	
	(0.0199)	(0.0413)	(0.0074)		(0.0116)	
250-500m	0.0190	0.0672**	-0.0052		0.0024	
	(0.0142)	(0.0315)	(0.0046)		(0.0080)	
R^2	0.00705	0.00478	0.00297		0.00033	
Observations	13,035	13,035	13,035		13,035	
Panel B. High-Skilled Offices						
0-250m	0.1133***	0.2024***	0.0267**	0.0111	0.0197	-0.0178
	(0.0412)	(0.0750)	(0.0130)	(0.0125)	(0.0228)	(0.0198)
250-500m	0.0172	0.0918	-0.0023	-0.0101	0.0071	-0.0072
	(0.0285)	(0.0597)	(0.0089)	(0.0096)	(0.0143)	(0.0145)
R^2	0.00765	0.00693	0.00384	0.00168	0.00079	0.00072
Observations	9,810	9,405	9,405	8,715	9,405	8,715
Panel C. Low-Skilled Offices						
0-250m	0.0658**	0.0273	0.0090	-0.0051	-0.0281*	-0.0343**
	(0.0330)	(0.0977)	(0.0112)	(0.0111)	(0.0168)	(0.0171)
250-500m	0.0365*	0.0295	-0.0007	-0.0114	-0.0031	-0.0409***
	(0.0220)	(0.0636)	(0.0080)	(0.0085)	(0.0111)	(0.0112)
R^2	0.00389	0.00011	0.00038	0.00121	0.00173	0.00860
Observations	11,610	11,430	11,430	10,830	11,430	10,830
Panel D. Local Services						
0-250m	0.0518*	0.0754	-0.0016	-0.0161**	0.0014	-0.0317***
	(0.0290)	(0.0485)	(0.0052)	(0.0061)	(0.0112)	(0.0099)
250-500m	-0.0039	-0.0030	-0.0066	-0.0062	-0.0052	-0.0147***
	(0.0159)	(0.0288)	(0.0042)	(0.0044)	(0.0072)	(0.0073)
R^2	0.00400	0.00259	0.00130	0.00516	0.00026	0.00722
Observations	11,865	11,685	11,685	11,505	11,685	11,505
Panel E. Non-Offices						
0-250m	0.0301	0.0646	0.0204**	0.0205**	0.0049	-0.0112
	(0.0342)	(0.0651)	(0.0095)	(0.0104)	(0.0169)	(0.0175)
250-500m	-0.0063	0.0024	-0.0044	0.0002	-0.0011	-0.0229*
	(0.0194)	(0.0389)	(0.0066)	(0.0079)	(0.0106)	(0.0121)
R^2	0.00081	0.00072	0.00344	0.00248	0.00009	0.00218
Observations	11,610	11,340	11,340	10,995	11,340	10,995

Notes: This table reports estimates of α_{250} and α_{500} in Equation (14) for different outcome variables indicated in the columns and using an alternative sample of cells, as described in Section 6.1. Standard errors clustered at the cell level are displayed in parentheses. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

D.2 Alternative Thresholds - New Buildings

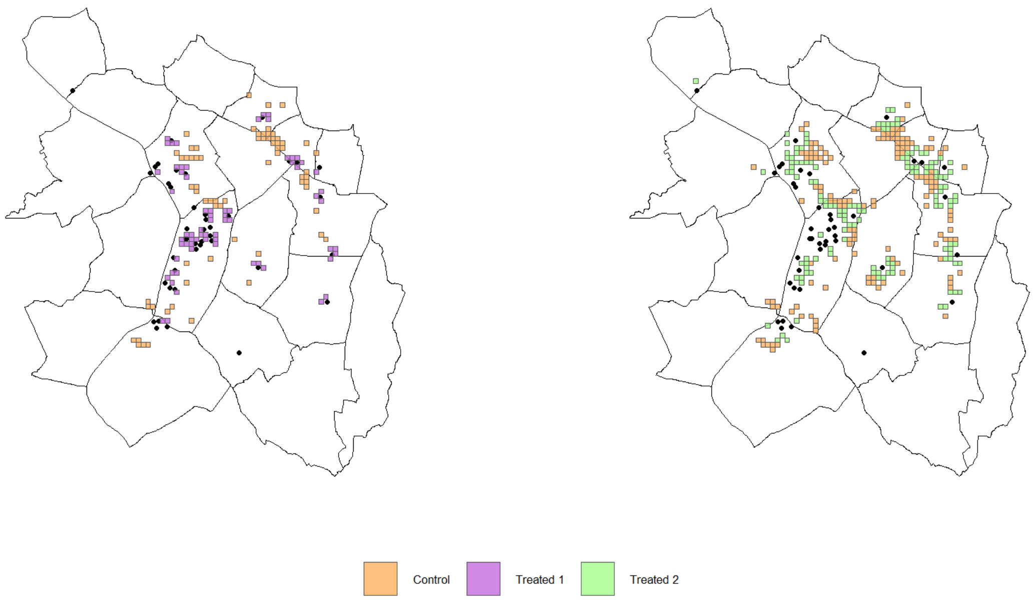
Table D2: Effects of New Commercial Buildings

Variable	Baseline	$n = 600$	$n = 400$	$sh = 15\%$	$sh = 35\%$
High-skilled Offices					
Log Estabs	0.0495 (0.0417)	0.0686 (0.0577)	0.0674 (0.0389)	0.0394 (0.0359)	0.0169 (0.0429)
Log Workers	0.2087 (0.0838)	0.1882 (0.1214)	0.2365 (0.0795)	0.1279 (0.077)	0.2024 (0.0889)
% College	0.0337 (0.0135)	0.0417 (0.0183)	0.0215 (0.0127)	0.0326 (0.0125)	0.0273 (0.0148)
Wage Premium	0.0609 (0.0252)	0.0801 (0.0388)	0.0636 (0.0226)	0.0701 (0.0242)	0.067 (0.028)
Low-skilled Offices					
Log Estabs	0.0677 (0.0373)	0.0715 (0.0562)	0.0486 (0.0341)	0.0523 (0.0331)	0.0483 (0.0394)
Log Workers	0.0743 (0.1112)	0.0641 (0.155)	-0.0066 (0.1017)	0.0864 (0.1018)	0.0181 (0.1197)
% College	0.004 (0.0128)	0.0078 (0.0181)	0.0014 (0.0121)	0.0075 (0.012)	0.0057 (0.0137)
Wage Premium	-0.0368 (0.0197)	-0.0201 (0.0293)	-0.0361 (0.019)	-0.0239 (0.0176)	-0.0343 (0.0212)
Local Services					
Log Estabs	0.077 (0.038)	0.1423 (0.0526)	0.0455 (0.037)	0.0595 (0.0289)	0.0696 (0.041)
Log Workers	0.1341 (0.0559)	0.2068 (0.0844)	0.1083 (0.0507)	0.1137 (0.048)	0.1489 (0.0605)
% College	-0.0067 (0.0067)	6e-04 (0.0101)	-0.0022 (0.0067)	-0.0074 (0.0065)	-0.0074 (0.0071)
Wage Premium	-0.0107 (0.0112)	0.001 (0.0164)	-0.0044 (0.0111)	-0.0097 (0.0107)	-0.0086 (0.0125)
Non-offices					
Log Estabs	0.0016 (0.0447)	0.0624 (0.0673)	-0.0028 (0.0382)	-0.0106 (0.0331)	-0.0154 (0.0417)
Log Workers	0.0798 (0.0779)	0.2517 (0.1101)	0.0956 (0.0725)	0.0411 (0.0584)	0.0173 (0.0672)
% College	0.0258 (0.0115)	0.0324 (0.0168)	0.0316 (0.011)	0.0238 (0.0112)	0.0277 (0.0129)
Wage Premium	0.0142 (0.0204)	0.0265 (0.0281)	0.0279 (0.0198)	0.0087 (0.0181)	0.0077 (0.0219)
Obs	7170	4425	7590	7935	6540

Notes: This table compares estimates of α_{250} in Equation (14) for various outcome variables and neighborhood samples based on alternative samples of new commercial buildings, as described in Section 6.2. n and sh account for the average employment and the average share of college-degree workers. In the baseline, $n = 500$ and $sh = 25\%$. Standard errors clustered at the cell level are displayed in parentheses.

D.3 Nearest Neighbor Matching using the Proximity Probability Measure

Figure D2: Empirical Analysis Setup



Notes: the figure depicts the design of the empirical analysis using nearest neighbor matching, as described in Section 6.3. The left figure depicts the pairing between first-ring cells with outer-ring cells and the right figure depicts the pairing between second-ring cells and outer-ring cells. The black dots indicate the location of new commercial buildings.

Table D3: Effects of New Commercial Buildings: Nearest Neighbor Matching (First-ring Cells)

	Log Estabs	Log Workers	% College		Wage Premium	
			All estabs	excl. new estabs	All estabs	excl. new estabs
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. High-Skilled Offices						
0-250m	0.0890*	0.1818**	0.0288**	0.0202	0.0527**	0.0132
	(0.0457)	(0.0867)	(0.0143)	(0.0133)	(0.0263)	(0.0215)
R^2	0.01257	0.01363	0.01238	0.00697	0.01282	0.00073
Observations	1,986	1,986	1,986	1,896	1,986	1,896
Panel B. Low-Skilled Offices						
0-250m	0.0614	0.0975	0.0024	-0.0167	-0.0229	-0.0349
	(0.0404)	(0.1201)	(0.0135)	(0.0130)	(0.0224)	(0.0225)
R^2	0.00688	0.00178	-0.00043	0.00345	0.00253	0.00735
Observations	1,986	1,986	1,986	1,836	1,986	1,836
Panel C. Local Services						
0-250m	0.0678*	0.1204*	-0.0025	-0.0129	-0.0105	-0.0350**
	(0.0396)	(0.0634)	(0.0081)	(0.0087)	(0.0134)	(0.0154)
R^2	0.01012	0.01389	-0.00023	0.00537	0.00133	0.01376
Observations	1,986	1,986	1,986	1,971	1,986	1,971
Panel D. Non-Offices						
0-250m	-0.0070	0.0977	0.0239*	0.0228	0.0114	0.0015
	(0.0540)	(0.0862)	(0.0136)	(0.0147)	(0.0220)	(0.0203)
R^2	-0.00041	0.00474	0.01024	0.00850	0.00036	-0.00051
Observations	1,986	1,986	1,986	1,911	1,986	1,911

Notes: This table reports estimates of new building effects for different outcome variables indicated in the columns using nearest neighbor matching, as described in Section 6.3. Standard errors clustered at the cell level are displayed in parentheses. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

Table D4: Effects of New Commercial Buildings: Nearest Neighbor Matching (Second-ring Cells)

	Log Estabs	Log Workers	% College		Wage Premium	
			All estabs	excl. new estabs	All estabs	excl. new estabs
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. High-Skilled Offices						
250-500m	0.0162	0.0559	0.0012	-0.0075	0.0126	-0.0055
	(0.0292)	(0.0696)	(0.0101)	(0.0111)	(0.0169)	(0.0169)
R^2	0.00019	0.00081	-0.00023	0.00044	0.00058	0.00009
Observations	4,043	4,043	4,043	3,668	4,043	3,668
Panel B. Low-Skilled Offices						
250-500m	0.0406	0.1019	-0.0126	-0.0219**	-0.0089	-0.0544***
	(0.0274)	(0.0766)	(0.0105)	(0.0106)	(0.0144)	(0.0140)
R^2	0.00322	0.00257	0.00194	0.00666	0.00026	0.01957
Observations	4,043	4,043	4,043	3,795	4,043	3,795
Panel C. Local Services						
250-500m	0.0133	-0.0020	-0.0095*	-0.0115**	-0.0158*	-0.0218**
	(0.0179)	(0.0338)	(0.0056)	(0.0055)	(0.0091)	(0.0094)
R^2	0.00054	-0.00024	0.00371	0.00616	0.00360	0.00709
Observations	4,043	4,043	4,043	3,968	4,043	3,968
Panel D. Non-Offices						
250-500m	-0.0341	-0.0722	0.0047	0.0128	-0.0042	-0.0311*
	(0.0264)	(0.0521)	(0.0093)	(0.0106)	(0.0141)	(0.0159)
R^2	0.00250	0.00258	0.00011	0.00204	-0.00013	0.00535
Observations	4,043	4,043	4,043	3,893	4,043	3,893

Notes: This table reports estimates of new building effects for different outcome variables indicated in the columns using nearest neighbor matching, as described in Section 6.3. Standard errors clustered at the cell level are displayed in parentheses. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

D.4 Excluding Second-ring Cells

Table D5: Effects of New Commercial Buildings: Excluding Second-ring Cells)

	Log Estabs (1)	Log Workers (2)	% College (3)	Wage Premium (4)
Panel A. High-Skilled Offices				
0-250m	0.0433	0.2089**	0.0351**	0.0620**
	(0.0419)	(0.0873)	(0.0139)	(0.0255)
R^2	0.00214	0.01173	0.01049	0.01361
Observations	5,145	5,145	5,145	5,145
Panel B. Low-Skilled Offices				
0-250m	0.0635*	0.1019	0.0022	-0.0412**
	(0.0374)	(0.1133)	(0.0129)	(0.0200)
R^2	0.00572	0.00171	-0.00014	0.00725
Observations	5,145	5,145	5,145	5,145
Panel C. Local Services				
0-250m	0.0733*	0.1260**	-0.0052	-0.0099
	(0.0383)	(0.0569)	(0.0070)	(0.0115)
R^2	0.01203	0.01282	0.00080	0.00104
Observations	5,145	5,145	5,145	5,145
Panel D. Non-Offices				
0-250m	0.0015	0.0772	0.0266**	0.0128
	(0.0453)	(0.0790)	(0.0117)	(0.0207)
R^2	-0.00019	0.00185	0.00811	0.00053
Observations	5,145	5,145	5,145	5,145

Notes: This table reports estimates of new building effects for different outcome variables indicated in the columns using only first-ring cells as the treatment group. Standard errors clustered at the cell level are displayed in parentheses. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

D.5 Continuous Treatment

For this estimation, I consider the following adaptation of Equation (14):

$$y_{c,t} = \alpha T_{c,t} + \Psi_c + \xi_{PP,t} + \mu_{D,t} + \upsilon'_{c,t} \quad , \tag{D.1}$$

where $T_{c,t} = (1-d) \times \mathbb{1}(Treated)$ is the new treatment variable based on the closest new building I use to separate neighborhoods into treatment and control groups (see Section 4.2). $\mathbb{1}(Treated)$ is an indicator of treatment, which also applies to cells previously used in the control group. Now, every neighborhood is treated to some extent, with the start of the treatment defined as the inauguration year of the "treatment" building. The variable d, in turn, represents the distance from this building. Note that, given this definition, $0 \leq T_{c,t} \leq 1$, with α expected to be positive.

Table D6: Effects of New Commercial Buildings: Continuous Treatment

	Log Estabs (1)	Log Workers (2)	% College (3)	Wage Premium (4)
Panel A. High-Skilled Offices				
dist	0.0854	0.3270^{***}	0.0395^{**}	0.1015^{***}
	(0.0611)	(0.1206)	(0.0182)	(0.0295)
R^2	0.89713	0.82653	0.61470	0.66977
Observations	7,170	7,170	7,170	7,170
Panel B. Low-Skilled Offices				
dist	0.0230	0.0820	-0.0181	-0.0168
	(0.0520)	(0.1472)	(0.0181)	(0.0293)
R^2	0.89867	0.74574	0.54145	0.62861
Observations	7,170	7,170	7,170	7,170
Panel C. Local Services				
dist	0.0748^{*}	0.0595	-0.0228^{**}	-0.0134
	(0.0452)	(0.0698)	(0.0093)	(0.0170)
R^2	0.93629	0.89501	0.72239	0.73889
Observations	7,170	7,170	7,170	7,170
Panel D. Non-Offices				
dist	0.0321	-0.0409	0.0105	-0.0183
	(0.0532)	(0.0958)	(0.0160)	(0.0267)
R^2	0.85306	0.79215	0.68072	0.71915
Observations	7,170	7,170	7,170	7,170

Notes: This table reports estimates of new building effects for different outcome variables indicated in the columns considering a continuous treatment variable according to Equation (D.1). Standard errors clustered at the cell level are displayed in parentheses. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

D.6 Only Cell and District-year Fixed-effects

This estimation ignores the propensity score and simply compares closer vs. further neighborhoods

$$y_{c,t} = \alpha_{250} D_{c,t,250} + \alpha_{500} D_{c,t,500} + \Psi_c + \mu_{D,t} + \upsilon''_{c,t} \quad . \tag{D.2}$$

Table D7: Effects of New Commercial Buildings: Only Cell and District-year Fixed-effects

	Log Estabs	Log Workers	% College	Wage Premium
	(1)	(2)	(3)	(4)
Panel A. High-Skilled Offices				
0-250m	0.0699*	0.1878**	0.0243*	0.0316
	(0.0393)	(0.0798)	(0.0137)	(0.0258)
250-500m	0.0394	0.0788	-0.0019	0.0009
	(0.0275)	(0.0632)	(0.0089)	(0.0147)
R^2	0.00439	0.00612	0.00354	0.00227
Observations	7,170	7,170	7,170	7,170
Panel B. Low-Skilled Offices				
0-250m	0.0709*	0.0359	0.0030	-0.0384**
	(0.0370)	(0.1093)	(0.0124)	(0.0195)
250-500m	0.0442*	0.0724	-0.0080	-0.0056
	(0.0263)	(0.0709)	(0.0094)	(0.0135)
R^2	0.00594	0.00092	0.00080	0.00398
Observations	7,170	7,170	7,170	7,170
Panel C. Local Services				
0-250m	0.0668*	0.1080**	-0.0116*	-0.0196*
	(0.0357)	(0.0537)	(0.0062)	(0.0108)
250-500m	-0.0116	-0.0194	-0.0113**	-0.0188**
	(0.0168)	(0.0302)	(0.0053)	(0.0083)
R^2	0.00927	0.00813	0.00570	0.00571
Observations	7,170	7,170	7,170	7,170
Panel D. Non-Offices				
0-250m	0.0086	0.0651	0.0175	0.0085
	(0.0423)	(0.0780)	(0.0108)	(0.0190)
250-500m	-0.0113	-0.0454	-0.0082	-0.0101
	(0.0230)	(0.0471)	(0.0083)	(0.0125)
R^2	0.00023	0.00227	0.00394	0.00080
Observations	7,170	7,170	7,170	7,170

Notes: This table reports estimates of new building effects for different outcome variables indicated in the columns according to Equation (D.2). Standard errors clustered at the cell level are displayed in parentheses. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

D.7 Clustering using Closest New Building

Table D8: Effects of New Commercial Buildings: Alternative Clustering

	Log Estabs	Log Workers	% College		Wage Premium	
			All estabs	excl. new estabs	All estabs	excl. new estabs
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. High-Skilled Offices						
0-250m	0.0495	0.2087^{***}	0.0337^{**}	0.0231	0.0609^{***}	0.0186
	(0.0451)	(0.0805)	(0.0146)	(0.0154)	(0.0198)	(0.0187)
250-500m	0.0260	0.0920	0.0043	-0.0016	0.0210	0.0088
	(0.0333)	(0.0797)	(0.0070)	(0.0081)	(0.0152)	(0.0147)
R^2	0.00206	0.00765	0.00620	0.00332	0.00873	0.00093
Observations	7,170	7,170	7,170	6,705	7,170	6,705
Panel B. Low-Skilled Offices						
0-250m	0.0677	0.0743	0.0040	-0.0128	-0.0368^{*}	-0.0464^{**}
	(0.0519)	(0.1086)	(0.0128)	(0.0116)	(0.0206)	(0.0196)
250-500m	0.0443	0.1035	-0.0073	-0.0174^{**}	-0.0041	-0.0515^{***}
	(0.0317)	(0.0753)	(0.0103)	(0.0085)	(0.0125)	(0.0104)
R^2	0.00563	0.00218	0.00074	0.00364	0.00367	0.01559
Observations	7,170	7,170	7,170	6,765	7,170	6,765
Panel C. Local Services						
0-250m	0.0770^{*}	0.1341^{**}	-0.0067	-0.0157^{**}	-0.0107	-0.0335^{**}
	(0.0421)	(0.0598)	(0.0062)	(0.0075)	(0.0121)	(0.0133)
250-500m	-0.0034	-0.0016	-0.0077	-0.0096	-0.0127	-0.0183^{*}
	(0.0237)	(0.0347)	(0.0061)	(0.0059)	(0.0083)	(0.0094)
R^2	0.01093	0.01076	0.00230	0.00640	0.00223	0.00909
Observations	7,170	7,170	7,170	7,080	7,170	7,080
Panel D. Non-Offices						
0-250m	0.0016	0.0798	0.0258^{**}	0.0258^{**}	0.0142	0.0030
	(0.0539)	(0.0863)	(0.0117)	(0.0120)	(0.0219)	(0.0165)
250-500m	-0.0109	-0.0292	-0.0032	0.0048	-0.0069	-0.0335^{***}
	(0.0221)	(0.0569)	(0.0083)	(0.0087)	(0.0108)	(0.0114)
R^2	8.25×10^{-5}	0.00219	0.00579	0.00420	0.00095	0.00555
Observations	7,170	7,170	7,170	6,945	7,170	6,945

Notes: This table replicates the sector results of Table 3 but clustering standard errors at the closest new building level, as explained above. Standard errors clustered at the cell level are displayed in parentheses. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.