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

Thiago Patto†

August 21, 2025

Click here for latest version

Abstract

I study how large commercial buildings impact different sectors locally. Using a ring approach and a matching procedure, I construct samples of neighborhoods with similar probabilities of observing a new building nearby. I estimate a local multiplier effect of one job in local services for every three additional jobs in high-skilled offices, resulting in an overall employment increase of 17.2%. High-skilled offices also experience increases in the share of college-educated workers and wages. These results are largely driven by newly created firms and the relocation of existing ones. Finally, I find suggestive evidence of increased industry diversity within high-skilled offices.

*I am grateful to Rodrigo Soares and Jorge De la Roca for their invaluable guidance and support. I would like to thank Enrico Moretti, Mariaflavia Harari, Daniel da Mata, Rogerio Santarrosa, David Card, Cristine Pinto, Nick Tsivanidis, and many seminar participants for helpful comments and suggestions. I acknowledge the São Paulo State Research Foundation (FAPESP) for their financial support (fellowship processes 2019/26148-0 and 2020/16601-7). All errors remain my own.

†Fundação Getúlio Vargas (FGV), São Paulo, *thiago.patto@fgv.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 São Paulo as a natural experiment, this article helps 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 differential effects. My sector classification is designed to distinguish between establishments that produce non-tradable goods and, therefore, are more dependent on local demand (local services)

[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 (2023), 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).

and those that produce tradable goods (high- and low-skilled offices).[3]

With 21.7 million inhabitants, São Paulo is one of the ten largest metropolitan areas globally and the largest in the Americas. It has almost doubled its population in 40 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 São Paulo municipal government (IPTU) and matched employer-employee data from the Ministry of Labor (RAIS). I use them for two purposes. First, I combine the datasets to identify recently inaugurated buildings that 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 estimate 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 the 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)

[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 Supplementary Appendix B gives more details.

[4] Source: OECD/European Commission (2020)

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

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 development. 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 perform a nearest-neighbor matching with replacement to construct the control groups used in the estimation.

The results show significant effects on local economic activity. Neighborhoods within 250 meters of a new building experience a 17.2% 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 18.5% and 31.2%, respectively. Considering the average employment one period prior to the treatment, it means that approximately one job is created in local services for every three jobs created in high-skilled offices. I also find results consistent with the idea that high-skilled office productivity is affected in nearby neighborhoods. There is a 4 percentage point differential increase in the share of college-educated workers and an 11.7% differential increase in the average wage premium paid by this sector.[6] Additional evidence indicates that most of these effects — on employment, wages, and worker composition — are not driven by incumbent establishments but rather by newly created ones and the internal relocation of existing establishments within São Paulo. The latter channel is particularly noteworthy, as it highlights an often-overlooked dimension of urban dynamics.

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

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

in the TFP of one sector in a 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 between local services and offices can be interpreted as a multiplier effect.

To shed more light on the spillover mechanism, I study the role of industry composition within high-skilled offices. I develop a broadly applicable metric that reflects changes in the composition of neighborhoods relative to nearby new buildings, drawing on definitions of industry similarity that have been proposed by the literature. I find that while neighborhoods with industries more heavily present in nearby new buildings (prior to the opening) experience stronger effects, less similar ones are also significantly affected. Moreover, productivity spillovers from new buildings seem to promote industry diversity.

I also explore the impacts of new buildings on other dimensions. I show that local services and offices expand at the extensive margin in treated neighborhoods, with a differential increase of 5% in the number of establishments, driven by high-skilled offices and local services. Regarding the supply and value of floor space, I find some evidence of an impact on nearby neighborhoods in the medium term.

Taken together, these findings indicate that productivity spillovers and local demand are fundamental drivers of spatial concentration. The opening of a new commercial building rapidly increases local employment and enhances the productivity of nearby high-skilled offices, especially high-wage firms. As a result, this sector expands its presence in these neighborhoods, which in turn increases demand for local services, leading to their expansion as well.

This paper adds to the literature on the distribution of economic activity within cities, particularly to studies on local agglomeration forces, pioneered by Arzaghi and Henderson (2008). Examples include cross-sectional analyses (Rosenthal and Strange, 2020; Liu et al., 2020, 2022) as well as studies employing quantitative spatial models to estimate agglomeration effects (Ahlfeldt et al., 2015; Heblich et al., 2020; Tsivanidis, 2023). I extend this literature by developing a methodology

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

based on local shocks that examines in detail sectoral differences in spatial distribution. In addition, I present new evidence on firm sorting and relocation within a city, consistent with the hypothesis that more productive firms are more sensitive to agglomeration effects.

In this sense, my findings are closely related and complementary to those of Baum-Snow et al. (2021), who estimate productivity externalities at a comparable spatial scale using a peer-effects model and firm revenue as a proxy for productivity. While their approach enables the estimation of structural parameters for counterfactual analysis, it is limited to high-skilled services. In contrast, this paper provides a broader view of the urban concentration process, using a different identification strategy based on the opening of new commercial buildings.

More broadly, this paper also relates to the literature on agglomeration economies (see Moretti, 2011; Combes and Gobillon, 2015, for related surveys). This body of work has emphasized various factors that shape the economy of cities, including firm sorting (Combes et al., 2012; Gaubert, 2018), worker sorting and the urban wage premium (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 revisit some of these topics at a finer spatial scale, documenting new patterns in the intra-urban distribution of economic activity.

Finally, this study builds on extensive literature exploring spatially distributed treatment effects. Two methodological strands of particular interest involve studies that examine the entry of large firms (Greenstone et al., 2010; Qian and Tan, 2021) and the construction of new buildings (Asquith et al., 2021; Pennington, 2021; Tsivanidis and Gechter, 2023). My empirical approach adapts insights from these papers — such as the combination of a ring-based design with matching — and proposes a new strategy that addresses potential endogeneity concerns.

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 design. Sections 5 and 6 present the results, and Section 7 provides the robustness checks. Finally, Section 8 concludes.

2 Data and Descriptive Evidence

2.1 Data Sources

RAIS covers Brazil's formal labor market, with a 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 last 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 São Paulo at the unit level.[8] It contains information related to tax collection, such as a property's designation (commercial or residential), land and construction area, land and floor space value (for tax purposes), construction year, and number of floors.

I use RAIS to create a panel of private establishments in the São Paulo Metropolitan Area between 2003 and 2017. Based on this sample, I also construct an annual panel of individuals linked to these workplaces[9] and use these data to estimate separately, for each year, wage premiums for each establishment according to expression $\log w_{it} = X_{it}\Gamma_t + \psi_{e(i)t} + \nu_{it}$, where $\log w_{it}$ is the log wage of individual i in year t, X_{it} is a group of controls, $\psi_{e(i)t}$ is an indicator for establishment e where individual i works in year t, and ν_{it} is an error term.[10] The estimated $\hat{\psi}_{e(i)t}$ represent the wage premia used in the analysis.

Then, observations in RAIS and IPTU are geocoded, with a success rate of about 97% and 96%, respectively. More than 85% 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

[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.

[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.

across years.

The analysis is performed at the neighborhood level, defined as a 200-meter square cell. For this purpose, I split São Paulo's territory into a contiguous set of cells and use the successfully geocoded observations to compute the variables of interest. While IPTU contains information on floor space supply and value, 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, described as follows:[11,12]

- Local services: includes retail, food, bank agencies, gyms and personal care
- High-skilled offices: includes information, 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 represent establishments of industries typically found in urban areas and produce goods and services relatively more tradable than local services. The term "office" is used because 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 large cities is especially pronounced and has increased 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 any of the three sectors

[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.

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 the extent to which 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 of the same firm can have different CNAE codes depending on their purpose. However, there are some exceptions where a single CNAE code is applied to 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 São Paulo Metropolitan Area that explores the occupation composition of establishments. Details are provided in the Supplementary Appendix B.[14]

2.3 Descriptive Evidence

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

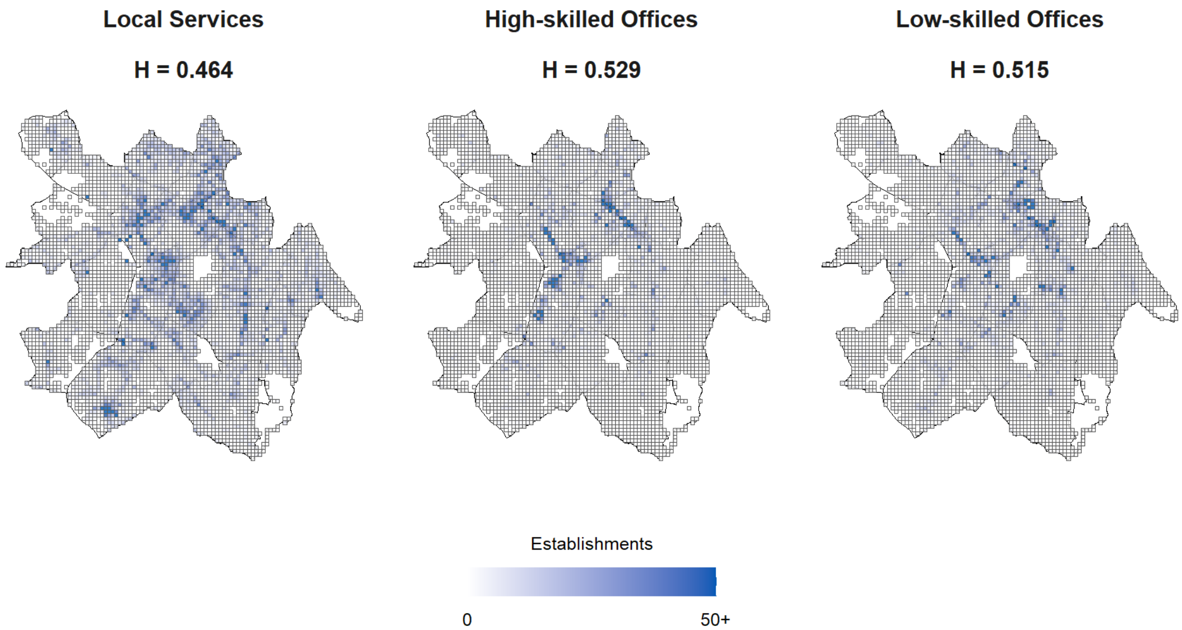
Related to the first point, Figure 1 shows the spatial distribution of establishments classified as local services, high-skilled offices, and low-skilled offices (as defined in the previous section) in 2010. The blue areas indicate where the establishments are located. I also display above the spatial Herfindahl index proposed by Guimaraes et al. (2011) for each sector.[15]

[13]The only group establishments excluded from the analysis is the public sector.

[14]Because this procedure requires a reasonable number of establishments, some may 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.

[15]This index is a more general version of the standard spatial Herfindahl index. It is given by $H = s'Ws$, where s is a vector containing the share of firms in each neighborhood for a given sector, and W is a weight matrix representing the neighbor structure. I define the term in row i and column j of the weight matrix as $w_{ij} = exp(-\frac{r_{ij}}{\bar{r}})$, where r_{ij} is the distance between the centroids of the neighborhoods i and j, and $\bar{r}$ is the average distance among all neighborhoods. Guimaraes et al. (2011) argue that their measure has the advantage of considering both the concentration level within

Figure 1. Spatial Distribution of Employment by Sector



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

There are significant differences in the concentration level of each of these sectors. Offices exhibit a more concentrated pattern, especially high-skilled offices. Most of the agglomeration takes place around three important avenues of São Paulo: Paulista (on the northeast side of the map), Brigadeiro Faria Lima, and Engenheiro Luis Carlos Berrini (in the center of the map). They represent, respectively, the old and new business centers of São Paulo and contain the highest employment densities in the city.[16] Local services spread relatively more throughout the city, but also have a higher presence where offices are located.

One potential factor behind this pattern is related to the nature of the goods produced by these

neighborhoods and the proximity between them. See their paper for a more detailed discussion.

[16]Historically, the first employment boom of São 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 Paulista Avenue. The traditional mansions were demolished to open up space for commercial buildings, which hosted the headquarters of many large companies. Brigadeiro Faria Lima Avenue began to gain relevance in the 1970s, and by the 1990s, it was already one of the densest employment areas. Today, it is considered the most important business center of São Paulo, especially for industries such as finance and technology. Employment in Luis Carlos Berrini Avenue followed the same trend as Faria Lima but with a few years' delay.

sectors. Businesses such as 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 technology 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.

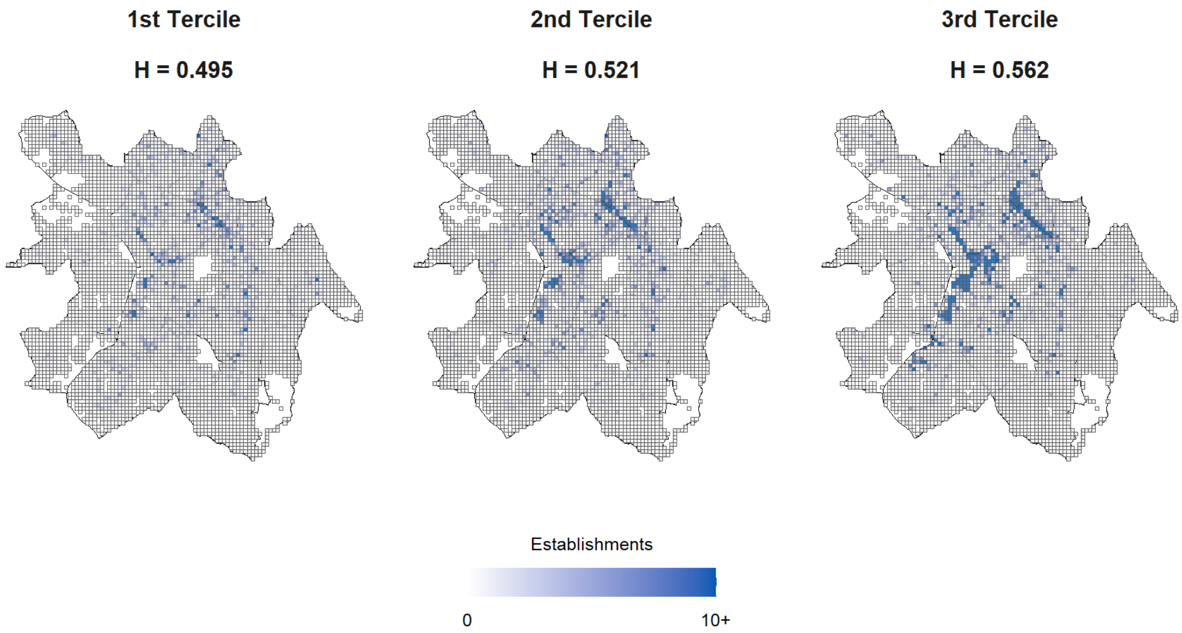
Regarding the second point, Figure 2 displays, for 2010, the distribution of establishments in high-skilled offices for different terciles of establishment wage premium. It is worth noting that establishments in the first tercile are more dispersed than in the third tercile, that is, the concentration level increases with the wage premium. To show that this fact is not entirely driven by industry composition, in Figure E1 I present the same analysis focusing on a 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 varies within and between sectors. In particular, the patterns observed for local services and offices seem to reflect different incentives to settle in these locations. Hence, to study the local effects of new commercial buildings, I propose a theory of firm location choice that rationalizes urban concentration as a combination of differences in local productivity and demand effects. Additionally, I discuss how firm heterogeneity can be relevant in this context.

3 Theoretical Framework

The first relevant question when studying the effects of new commercial buildings on local economic activity is how this influence is exerted. I consider three possible channels. First, new buildings can represent a shift in local demand as workers and firms that occupy the newly

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



Notes: This 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.

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 may stem, for example, from the spread of knowledge produced by newly arrived firms (Jaffe et al., 1993; Atkin et al., 2022) or improvements in firm-worker matching (Dauth et al., 2022).

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. The analysis focuses on changes in local productivity, as I argue that this channel is central to understanding the process of urban concentration. I assume at first that local productivity is exogenous, abstracting from productivity spillovers, and examine how an increase in this variable in a specific sector and neighborhood affects the local equilibrium. Given that my empirical analysis is based on small-scale shocks — unlikely to represent a significant departure from the city's prior equilibrium but with potential sizable local effects — I also assume that general equilibrium effects are negligible. These approximations allow me to derive testable predictions through simple analytical solutions, which can be compared with the empirical findings.

Next, I introduce productivity spillovers and discuss how a shift in available floor space in one neighborhood — such as the construction of a large commercial building — can affect surrounding neighborhoods. I consider a functional form with decaying effects and show how an increase in employment in the neighborhood that receives the new building generates differential effects across nearby neighborhoods, depending on their distance from the shock.

Finally, I consider an environment with firm heterogeneity and discuss the conditions under which we would observe firm sorting after a productivity shock. The model derivations are detailed in the Supplementary Appendix C.

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.[17] 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 therefore 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, that is, $w_{e,s,n} = w_{s,n}\ \forall e$.

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 throughout the city and to

[17]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.

the broader economy at a common price p_s, which is treated as fixed. In contrast, local services sell their goods locally at a price $p_{LS,n}$. For simplicity, I assume that this good is entirely consumed by workers within 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

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

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

$$\ell_n = B_n \left(\frac{w}{p_{LS,n}^{\delta}} \right)^{\varepsilon}. \tag{2}$$

In this expression, ε represents the elasticity of labor with respect to wages and B_n is the firm commuter market access (Tsivanidis, 2023). This term reflects how easily firms in a neighborhood can attract workers and accounts for differences in accessibility and amenities. Supplementary Appendix C shows a possible microfoundation 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 (1), the wage set by a firm in sector s that locates in n is given by

$$w_{s,n}=\left[\frac{1-\beta}{\varepsilon+1}\frac{\varepsilon}{\beta}\right]\beta^{\frac{1}{1-\beta}}\cdot\begin{cases}p_s^{\frac{1}{1-\beta}}\left(\frac{A_{s.n}}{r_n^{\beta}}\right)^{\frac{1}{1-\beta}} & \text{if } s\neq LS\\ \left(\frac{p_{LS,n}}{r_n^{\beta}}\right)^{\frac{1}{1-\beta}} & \text{if } s=LS\ .\end{cases} \tag{3}$$

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 (1) 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:

$$E_{s,n}=\left(\frac{A_{s,n}^{\chi}B_n}{r_n^{\chi\beta}p_{LS,n}^{\delta\varepsilon}}\right)^{\eta}\frac{\overline{E}_s}{\Phi_s}, \tag{4}$$

where $\Phi_s=\sum_i(A_{s,i}^{\chi}B_i/r_i^{\chi\beta}p_{LS,i}^{\delta\varepsilon})^{\eta}$ and $\chi=(1+\varepsilon)/(1-\beta)$. The term Φ_s shows how the neighborhoods are connected, thus capturing the general equilibrium effects of the model.

Equation (4) tells us that the relative sectoral presence in a neighborhood depends positively on local TFP and firm commuter market access, and negatively on rent and local service 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.[18]

For local services, the same procedure yields

$$E_{LS,n}=\left(\frac{p_{LS,n}^{\chi-\delta\varepsilon}B_n}{r_n^{\chi\beta}}\right)^{\eta}\frac{\overline{E}_{LS}}{\Phi_{LS}}, \tag{5}$$

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.

[18] 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.

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:

$$Y_{LS,n}E_{LS,n} = \frac{\delta}{p_{LS,n}} \sum_{s} w_{s,n}\ell_{s,n}E_{s,n}, \tag{6}$$

where $Y_{LS,n}$ is the amount of non-tradable goods produced in neighborhood n. Using (4), (5), and the first order conditions of (1) to solve for $Y_{LS,n}$, $w_{s,n}$ and $\ell_{s,n}$, $p_{LS,n}$ can be expressed as

$$p_{LS,n} = \left[\left(\frac{1}{\chi/\delta\varepsilon - 1}\right)\frac{\Phi_{LS}}{\overline{E}_{LS}} \sum_{s\neq LS} \frac{\overline{E}_s p_s^{\chi}}{\Phi_s} \cdot A_{s,n}^{\chi(1+\eta)}\right]^{\frac{1}{\chi(1+\eta)}}. \tag{7}$$

Equation (7) 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 the supply and demand for floor space in each neighborhood. To fix ideas, I assume a fixed supply $\overline{T}_n$, which can be interpreted as a short-term approximation.[19]

$$\overline{T}_n = \sum_{s} f_{s,n}E_{s,n}. \tag{8}$$

To derive an expression for r_n, I use again (4), (5) and the first order conditions of (1), together with Equation (7), to get

$$r_n = \left[K\frac{\overline{E}_{LS}}{\Phi_{LS}} \cdot \frac{B_n^{1+\eta}}{\overline{T}_n} \cdot p_{LS,n}^{(\chi-\delta\varepsilon)(1+\eta)}\right]^{\frac{1}{1+\beta\chi(1+\eta)}}, \tag{9}$$

where $K = \frac{1}{\delta}\left(\frac{\chi}{\varepsilon}\right)^{1-\varepsilon}\beta^{\chi-\varepsilon}$. Since $\chi - \delta\varepsilon > 0$, it is clear from Equations (7) and (9) that higher productivity is associated with higher prices of local services and floor space, as expected.[20]

[19] In Section 5.4, I analyze how new commercial buildings affect the supply and price of floor space.

[20] An alternative approach is to assume that there is a competitive construction sector that combines land and capital to produce floor space, and a fixed supply of available land. If this sector uses a Cobb-Douglas technology in which the share of expenditures on land is α, the floor space supply curve in neighborhood n can be written as $\overline{S}_n r_n^{\frac{1-\alpha}{\alpha}}$, where $\overline{S}_n$ is the total available land in neighborhood n. In this case, Equation (9) would be slightly modified, with

To understand how changes in local productivity can affect 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 $\frac{\partial \Phi_{s'}}{\partial A_{s,n}} \approx 0 \; \forall \; s'$, 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 A ■

After establishing from Equations (7) and (9) that the prices of local services and floor space increase due to an increase in the local TFP of one sector, it is straightforward from Equation (4) that the presence of all office sectors $s' \neq s$ in neighborhood n is 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 affected due to the 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 affect all sectors equally. Local productivity, in turn, is sector-specific, and sectors with higher productivity (i.e., higher $A_{s,n}$) will have more presence in the neighborhood.

Proposition 2 *Consider an increase in $A_{s,n}$. Assuming that $\frac{\partial \Phi_{s'}}{\partial A_{s,n}} \approx 0 \; \forall \; s'$, nominal wages in sector s and local services increase, whereas it decreases for all other 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 A ■

---

the outer exponent equal to $\frac{1}{1/\alpha + \beta\chi(1+\eta)}$ and $\overline{T}_n$ replaced by $\overline{S}_n$, but the conclusions of this section would remain unchanged.

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, that is, if the dispersion of idiosyncratic preferences is low, rent prices will be more sensitive to changes in the price of local services. Because real wages set by firms depend 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 $p_{LS,n}$ to changes in $A_{s,n}$, making 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 A discusses this matter in more detail.

Proposition 2 is particularly useful in the context of this paper, as it helps to distinguish a productivity shock from an amenity shock. Equations (3), (7) and (9) show that an increase in B_n would lead to a decrease in wages, in contrast to an increase in $A_{s,n}$.

3.2 Productivity Spillovers

Having examined the effects of local productivity on local equilibrium outcomes, I now consider spillover effects. Building on Ahlfeldt et al. (2015), I assume that $A_{s,n}$ has the following functional form:

$$A_{s,n} = a_{s,n} \Upsilon_n^{\lambda_s} , \quad \Upsilon \equiv \sum_i e^{-\tau_{n,i}} L_i , \tag{10}$$

where $\tau_{n,i}$ is a measure of distance between neighborhoods n and i, and $L_i = \sum_s \ell_{s,i} E_{s,i}$ is the total employment in neighborhood i. Local productivity for each sector is thus modeled as a combination of location fundamentals ($a_{s,n}$) and productivity spillovers (Υ_n) that decay with distance. Importantly, the strength of these spillovers may vary across sectors, depending on the magnitude of λ_s. [21]

While the introduction of spillovers may amplify the effects of an exogenous shift in local productivity, it does not alter the qualitative content of Propositions 1 and 2. However, the spatially decaying nature of spillovers provides a valuable source of variation that I exploit in this paper. The empirical analysis based on the opening of large commercial buildings is grounded on the notion that a sharp increase in floor space (and consequently, in employment) in one neighborhood can generate a significant productivity shock in nearby neighborhoods via spillover effects. When a large building opens in neighborhood n, some sectors in neighborhood n' may experience a sudden increase in $\Delta A_{sn'}$, with the magnitude of this increase declining with distance from n. By comparing neighborhoods at varying distances from new buildings, I am able to identify differential effects of local productivity shocks on the spatial distribution of economic activity. Moreover, the predictions obtained in the previous section provide a framework for interpreting the empirical findings.

3.3 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-neighborhood 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

[21] One might consider modeling productivity spillovers for a given sector as a function of sectoral employment instead of total employment. In Section 6, I investigate to what extent the industry composition of new buildings is connected to the main results that I report and if this composition affects industry sorting.

$$\frac{\partial^2 A_{esn}(\varphi_e, A_{sn})}{\partial \varphi_e \partial A_{sn}} > 0.$$

In this setting, it follows from Equation (11) that firms with higher φ_e are more likely to locate in highly productive neighborhoods. Furthermore, increases in local TFP amplify spatial sorting: as neighborhoods become more productive for particular sectors, they attract a greater share of more productive firms — possibly at the expense of less productive ones within the same sector.

4 Empirical Strategy

This section outlines the empirical approach of this paper, which investigates the effects of new large commercial buildings on economic activity. 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 build. From a broader perspective, new developments 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, further increasing uncertainty regarding the timing and location of the inauguration.

My empirical strategy aligns with these arguments by estimating the effects of new buildings from slight variations in distance and time, but refines this approach. I use a propensity score model to predict which cells are more likely to observe a new building in its vicinity and leverage this information to construct a control group through a matching procedure. In doing so, I compare neighborhoods with a similar likelihood of observing a new building nearby.[22] 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, discarding those with specific purposes that do not fit into this context, such as hotels, schools and religious temples.[23] Then, I merge 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.[24]

Based on this new information, I establish the following criteria to select the "treatment" buildings: i) inauguration year between 2006 and 2013, ii) at least 500 workers on average starting from the inauguration year, and iii) 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 E2 shows that the occupation rate of these buildings accelerates rapidly after the inauguration.[25]

[22] 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, neighborhoods located in central areas of São 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.

[23] 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 reclassifications, making it difficult to observe them consistently in the data over the years.

[24] The IPTU data are not suitable on their own to identify shocks in local employment for two reasons. First, they do not provide good information about the buildings' inauguration year. 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 developments typically appear in the IPTU data about two years after their actual completion. Secondly, some buildings may have a negligible employment impact if firms do not occupy them.

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

Table 1. Summary Statistics: New Commercial Buildings

	Median	Mean	Std. Deviation	Min	Max
Total Land Area (m^2)	2,857	6,612.6	10,917.3	750	54,082
Occupied Land Area (m^2)	2,100	3,118.3	2,786	573	11,481
Built-Area-Ratio	9.7	11.1	5.4	5.4	33.8
Establishments	13.5	22.6	32.4	1	175.6
Workers	773.1	1,280.4	1,764.7	511.5	11,965.9
% College	69.1	68.4	21.1	26.4	98.6
% of High-skilled Office workers	33.5	37.6	29.1	0	99.3
% of Low-skilled Office workers	5.3	15.7	24	0	94.6
% of Local Services workers	3.2	12.3	21.7	0	100
% of Non-office workers	23.7	34.2	30.5	0	99.8

Notes: This 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.

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 large company settles, but in general multiple establishments occupy the newly available space.[26]

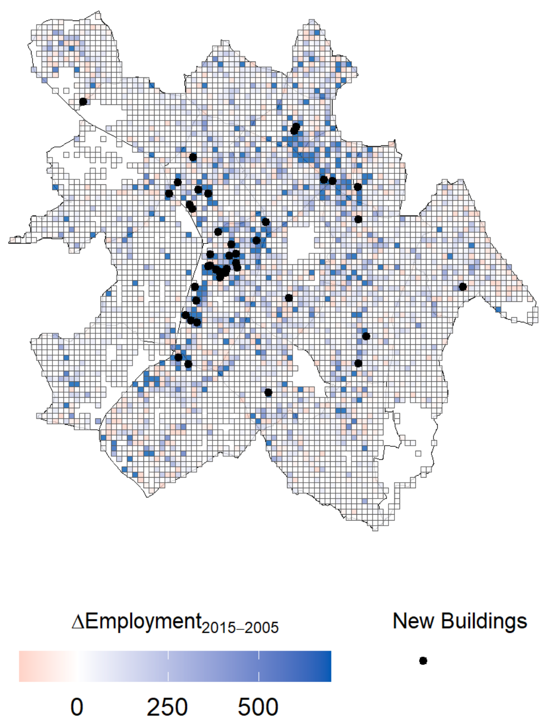
Figure 3 displays the location of the buildings alongside changes in neighborhood employment between 2005 and 2015. It is worth noting that, during this period, employment grew more rapidly in areas surrounding these buildings, reinforcing the possibility that their location reflects local employment trends. Another aspect to note is that many neighborhoods observe multiple new buildings nearby. These patterns present empirical challenges, which I address in the next section.

A comparison with Figure 1 reveals that these buildings are predominantly located in neighbor-

sector.

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

Figure 3. Location of New Commercial Buildings in São Paulo



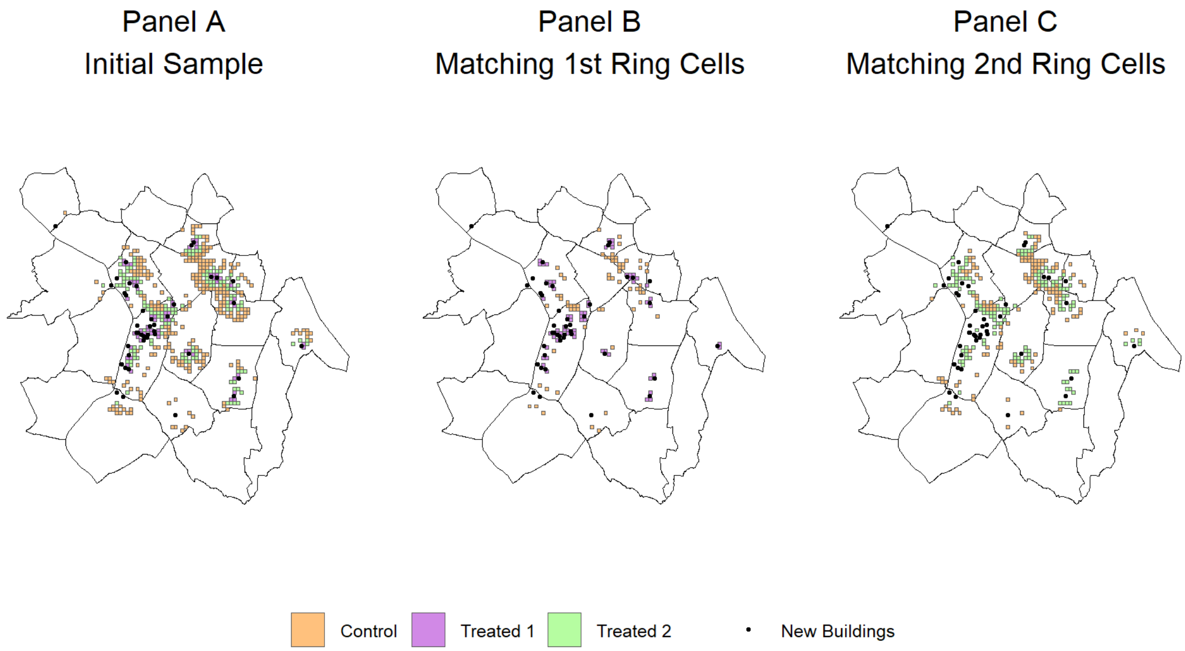
Notes: This figure shows in black dots the location of the new commercial buildings considered in the analysis, together with cell variation in employment between 2005 and 2015.

hoods with high employment density, near the city's major business corridors mentioned in Section 2.3—Paulista, Brigadeiro Faria Lima and Engenheiro Luis Carlos Berrini. This is unsurprising, as developers typically expand commercial floor space in areas with strong demand.

In order to validate the sample obtained, I use Google Maps to find the buildings and take screenshots of them. Some of these images are shown in Figure E3. 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 E4 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.[27]

[27]Google Maps imagery showing buildings under prolonged construction may raise concerns about anticipatory

Figure 4. Empirical Analysis Setup



Note: This figure depicts the design of the empirical analysis. Panel A shows the initial sample of treated and control cells. The black dots represent the new commercial buildings, and the solid lines represent district borders of São Paulo. Panels B and C present the results of the matching procedure using the Proximity Probability Score for each treated group. See the text for more details.

4.2 Selecting Treated and Control Neighborhoods, Defining Treatment

After selecting the new buildings to serve as shocks, I construct the sample of neighborhoods for the empirical analysis. Departing from the grid that covers São Paulo, I first select cells whose centroid lies within 1 km of at least one new building. Then, I exclude cells that received any of the 43 new buildings, as the goal is to explore shock external to the local neighborhood equilibrium. Next, I remove all establishments that ever settled in one of the new buildings, to avoid mechanical effects from firms relocating to these developments. Finally, to focus on changes at the intensive margin and ensure comparability of treatment effects across sectors, I restrict the analysis to cells containing at least one worker in each sector in all periods. This procedure yields an initial sample of 456 cells observed between 2003 and 2017.[28]

The next step is to define treatment. A fundamental feature of my empirical setting, illustrated

effects. However, the event-study estimates show no evidence of such dynamics in the variables analyzed.

[28] As a robustness check, I provide alternative results relaxing the last restriction, reported in Table F1. Supplementary Appendix F gives more details.

in 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 developments. 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;
- Treated Group 2 - second-ring cells: closest new building is within 250 to 500 meters;
- Control Group - outer-ring cells: closest new building is within 500 to 1000 meters.

Moreover, I define the first treatment period as the year I observe the first building inauguration within the distance bin associated with the cell. Figure 4 panel A shows the spatial distribution of the 456 cells with the proposed classification. The black dots represent the new commercial buildings and the solid lines represent the district borders of São Paulo.[29]

These definitions imply that my empirical design is not meant to identify the effects of one commercial building, but rather the differential exposure to increases in employment density in the vicinity due to these constructions. While this approach delivers estimates with a less intuitive interpretation, it is worth considering that new buildings are fairly heterogeneous and may not be fully occupied in the very short term (see Table 1 and Figure E2). Thus, the interpretation would not be straightforward even in an ideal scenario in which buildings are sufficiently distant from each other.

Additionally, it is important to emphasize that terms such as "treated group" and "control group" are used primarily for expositional clarity and to align with the terminology commonly used in the literature on the econometric methods employed in this paper. However, this does not imply that new buildings have no impact on control cells. Strictly speaking, my empirical strategy identifies only differential effects across neighborhoods with varying levels of exposure.

[29]To clarify how treatment start is defined, consider a cell located near 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.

To improve the comparability between treated and control cells, my research design also combines the "ring" method with a matching procedure to select a subset of control cells. Specifically, I estimate the probability that a cell is exposed to a new building in its vicinity using 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 of one of these buildings. It includes information on employment, wages, demography, and transportation access from different sources. Many of these variables are included in level and variation prior to 2006. Table E1 shows the complete list.

Equation (12) is estimated using Lasso. For this purpose, I use an extended sample that contains the cells described in the previous section, cells that received a new building, and a group of 'peripheral' cells located within 250 meters of a treated/control cell. The reason for including the last group will be explained below. Table E2 presents the parameters obtained for the chosen lambda, and Figure E5 exhibits the spatial distribution of the fitted values.[30]

My objective is to select outer-ring cells that, on average, have a probability of being treated similarly to that of first- and second-ring cells. Hence, instead of using the fitted values directly for the matching, I first compute the average propensity score in a 250-meter radius circumference drawn from the centroid of each cell. I interpret these numbers as measures related to the probability of being close to a new building, henceforth Proximity Probability Score (PPS). Since this method requires a propensity score for all neighbors of each cell in the initial sample, I need to include the peripheral cells in the estimation. Figure E6 presents PPS histograms for treated and control groups.

Finally, I use the PPS to perform a nearest-neighbor matching with replacement for each treated group. The final samples are depicted in Panels B and C of Figure 4. The first sample contains 126 cells (63 first-ring, 63 outer-ring), and the second sample contains 276 cells (138 second-ring,

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

Table 2. Baseline Summary Statistics: Treated and Control Groups

Variable	Treated 1	Control 1	t-test	Treated 2	Control 2	t-test
Workers						
High-skilled Offices	243.0 (382.4)	305.7 (705.2)	[0.284]	159.4 (390.4)	205.3 (556.6)	[0.170]
Low-skilled Offices	162.7 (440.8)	388.9 (1111.1)	[0.010]	225.8 (722.5)	215.6 (776.6)	[0.845]
Local Services	208.2 (215.3)	249.0 (541.2)	[0.337]	171.1 (233.4)	180.3 (386.7)	[0.680]
Non-offices	253.3 (337.5)	220.1 (387.6)	[0.375]	194.8 (297.1)	165.9 (287.3)	[0.156]
Establishments						
High-skilled Offices	14.8 (14.8)	14.6 (14.6)	[0.905]	10.7 (11.8)	9.9 (11.8)	[0.375]
Low-skilled Offices	14.4 (18.0)	14.4 (13.2)	[0.982]	12.6 (12.3)	12.0 (12.7)	[0.516]
Local Services	20.6 (14.4)	24.1 (33.1)	[0.184]	20.3 (15.4)	19.8 (25.0)	[0.711]
Non-offices	13.4 (10.7)	12.7 (10.5)	[0.510]	12.0 (9.2)	10.2 (8.4)	[0.003]
Wages						
High-skilled Offices	4937.6 (4076.7)	4947.0 (5162.9)	[0.984]	4061.8 (3227.0)	3969.7 (3946.9)	[0.713]
Low-skilled Offices	3325.2 (3062.4)	2351.7 (1277.8)	[0.000]	2492.8 (2043.8)	2167.1 (1306.3)	[0.006]
Local Services	2926.9 (2485.3)	2574.4 (2072.0)	[0.135]	2432.9 (1625.3)	2240.6 (1608.3)	[0.087]
Non-offices	4961.9 (3711.1)	4261.1 (2536.6)	[0.033]	4092.7 (2908.0)	3899.3 (2913.6)	[0.339]
% College						
High-skilled Offices	58.8 (24.0)	58.9 (21.1)	[0.974]	55.8 (21.6)	54.9 (22.5)	[0.572]
Low-skilled Offices	43.5 (22.1)	35.5 (19.1)	[0.000]	37.1 (20.9)	34.1 (19.3)	[0.030]
Local Services	29.7 (19.6)	25.8 (20.4)	[0.053]	25.0 (16.7)	24.0 (16.9)	[0.351]
Non-offices	39.7 (21.2)	42.9 (20.3)	[0.134]	41.0 (21.6)	40.3 (20.5)	[0.627]
Observations	63	63		138	138	

Notes: This table presents baseline summary statistics of treated and control groups using pre-treatment observations, i.e., prior to 2006. Standard deviations of variables appear in parentheses and p-values for differences of means appear in square brackets. Columns (1) and (2) show the mean and standard deviations for treated and control cells in the first sample, respectively, and Columns (4) and (5) do the same for the second sample. Columns (3 and (6) show the p-value of the t-tests of the difference in means in each case. Average wages are in 2017 reais.

138 outer-ring). In Table 2, I present baseline summary statistics for each sample and a balance test for the main outcomes analyzed in this study. Treated cells tend to have slightly higher wages for some sectors, but overall, the groups are reasonably similar on average. A comparison of these numbers with those of the initial sample, depicted in Table F3, reveals that the matching significantly enhances the similarity of the treated and control groups.

4.3 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

$$y_{c,t} = \sum_{k=-4}^{5} \alpha_{k,r} D_{c,k,t,r} + \Psi_c + \mu_{d,t} + \epsilon_{c,t} \quad . \tag{13}$$

In this expression, the subscript r alludes to one of the treated groups (first- or second-ring cells), and the subscript k represents event periods relative to t. The treatment variable $D_{c,k,t,r}$ takes the value 1 if cell c is in the treated group r and if the difference between year t and the year of treatment adoption is k. Thus, $\alpha_{k,r}$ represents the average effect of being differentially exposed to new buildings k periods from the start of treatment for group d. The specification also includes cell fixed effects Ψ_c and an interaction of district (d) and time indicators $\mu_{d,t}$.[31]

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:

$$y_{c,t} = \alpha_r D_{c,t,r} + \Psi_c + \mu_{d,t} + \upsilon_{c,t} \quad , \tag{14}$$

where $D_{c,t,r}$ is an indicator of whether cell c is treated in period t, and α_r is the treatment effect of group r now averaged over time.

I estimate Equations (13) and (14) separately for each sample using the two-stage procedure proposed by Gardner et al. (2024), which consists of regressing $y_{c,t}$ on Ψ_c and $\mu_{d,t}$ using only the untreated observations and then regressing the adjusted outcomes $y_{c,t} - \hat{\Psi}_c + \hat{\mu}_{d,t}$ on $D_{c,t,r}$ (or $D_{c,k,t,r}$ when estimating the event-study equation). Under parallel trends and no anticipation assumptions, this approach delivers estimates robust to heterogeneous treatment effects over cells

[31] The coefficients related to $k = -1$ are normalized to zero. If k is lower than -4 or greater than 5, I consider that $D_{c,-4,t,r} = 1$ and $D_{c,5,t,r} = 1$, respectively, i.e., those event periods are "binned".

and periods. Standard errors are clustered at the cell level.[32,33]

5 Results

I now present estimates of the effects of new commercial buildings 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 delve into the interpretation of the results.

5.1 Event Study

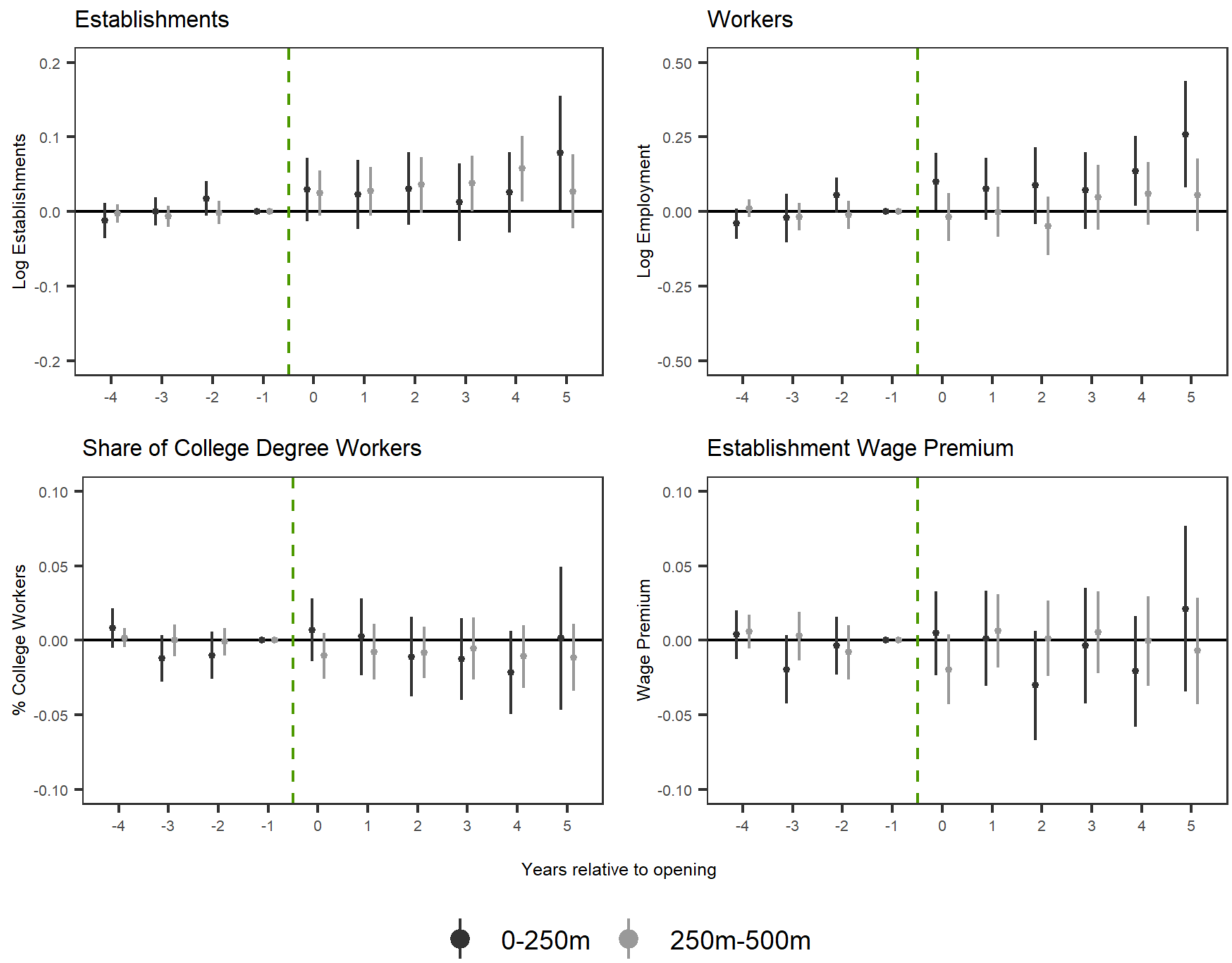
I begin by analyzing the impact of new commercial buildings on aggregate outcomes. The upper left panel of Figure 5 displays estimates from Equation (13) for the number of establishments. Both first- and second-ring cells appear to be differentially affected, though the standard errors are too large to draw a definitive conclusion. The upper right panel shows that employment follows a similar pattern, especially for first-ring cells. Importantly, the increase in point estimates persists over the period of analysis, and there is no sign of pre-trends prior to the beginning of treatment. The lower left and right panels of Figure 5 present 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. If anything, they suggest a negative differential effect on treated cells.

While the upper panels point to a differential increase in the level of economic activity, the lower panels indicate stability, or at least a small differential decrease, in local productivity. According to the theoretical discussion in Section 3, we should expect both variables to be positively affected

[32] Borusyak et al. (2021), Liu et al. (2022), and Wooldridge (2021) propose similar estimators with minor differences between them. One particular feature that makes this method more suitable for 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.

[33] As a robustness check, Table F9 reports estimates with standard errors clustered at the nearest new building. The results remain largely unchanged.

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

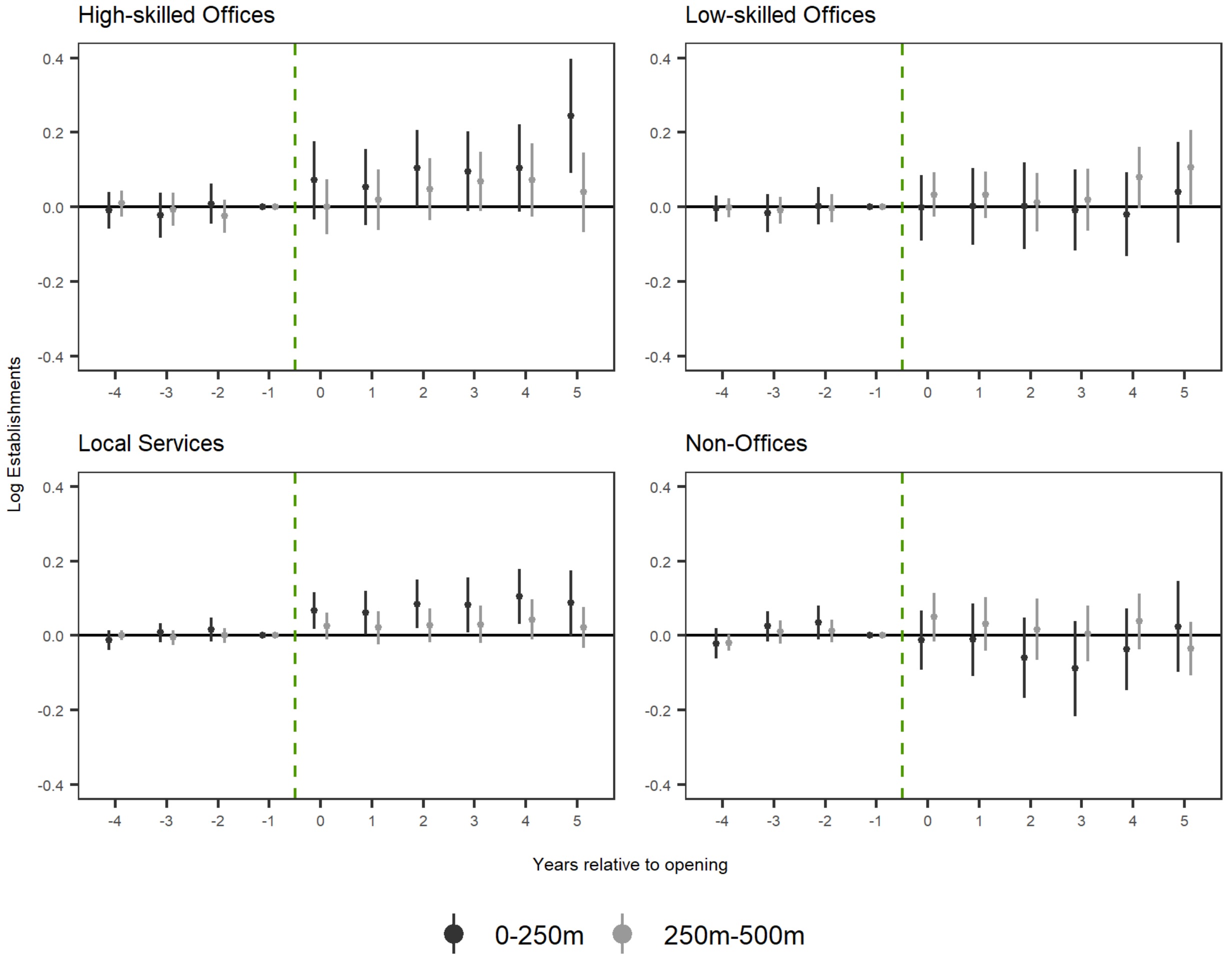


Note: This figure plots coefficients from running Equation (13) on the log number of establishments (upper left panel), the log employment (upper right panel), the share of workers with college degree (lower left panel) and the wage premium (lower right panel). The bars indicates the 95% confidence interval,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

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 might not be evident 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 in first-ring cells that takes effect in the first period after treatment initiation. Although the null hypothesis cannot be rejected at the 5% significance level, the results also suggest an increase in the number of high-skilled offices. Second-ring cells appear to be affected as well in later periods,

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



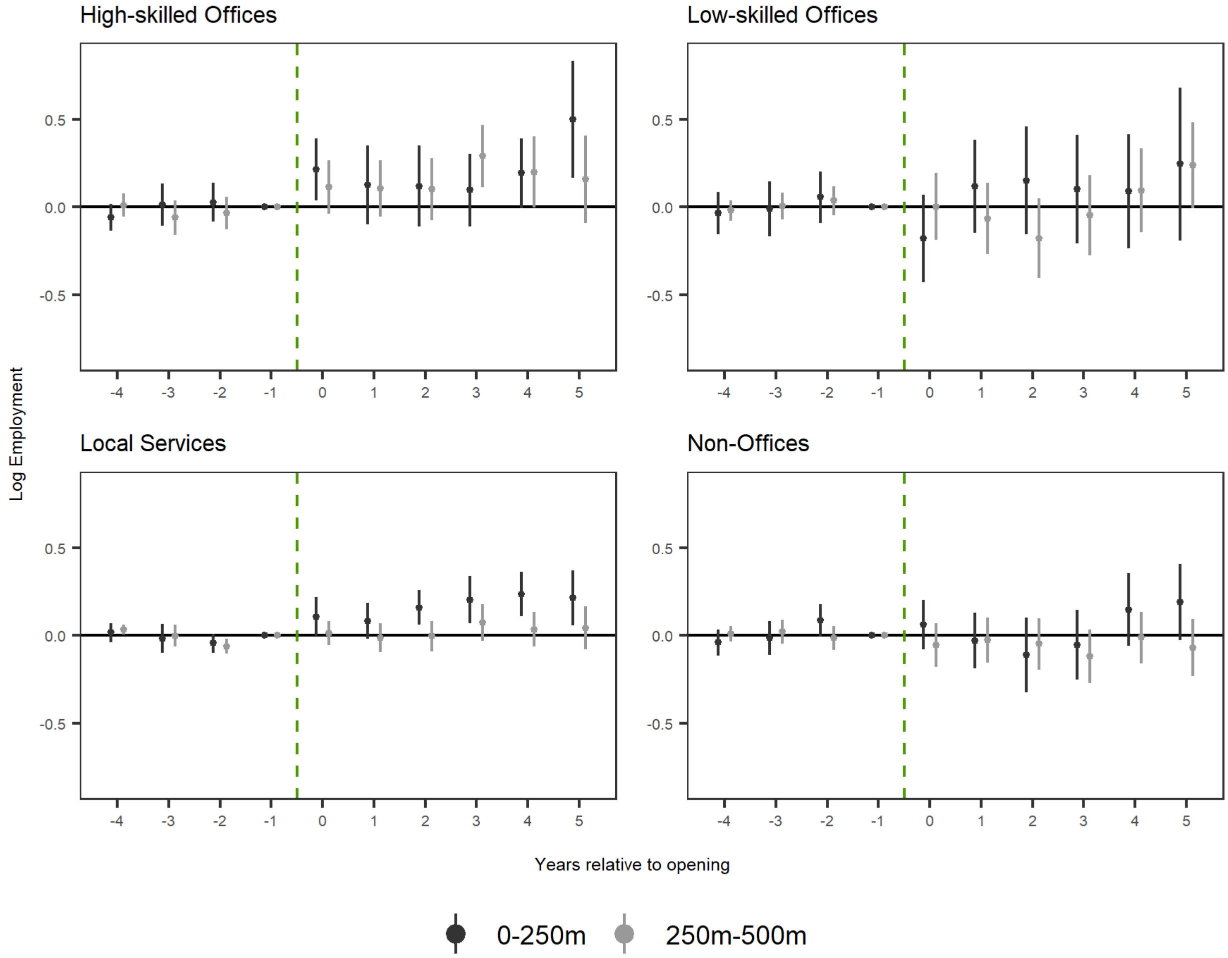
Note: This figure plots coefficients from running Equation (13) on the log number of establishments for different sectors. The definition of each sector is described in Section 2.2. The bars indicates the 95% confidence interval, where standard errors are clustered at the cell level.

though to a lesser extent.

Figure 7 displays the results for employment. Again, there is a differential (and persistent) increase in local services and high-skilled offices in first-ring cells, as well as in high-skilled offices in second-ring cells. This result can be interpreted, through the lens of the model, as a manifestation of the two types of agglomeration forces considered in this paper. The increased presence of offices would indicate that new buildings boost local productivity through spillover effects, whereas the rise in local services points to higher local demand for non-tradable goods.

To further examine this possibility, I estimate how wages and the share of college-educated

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

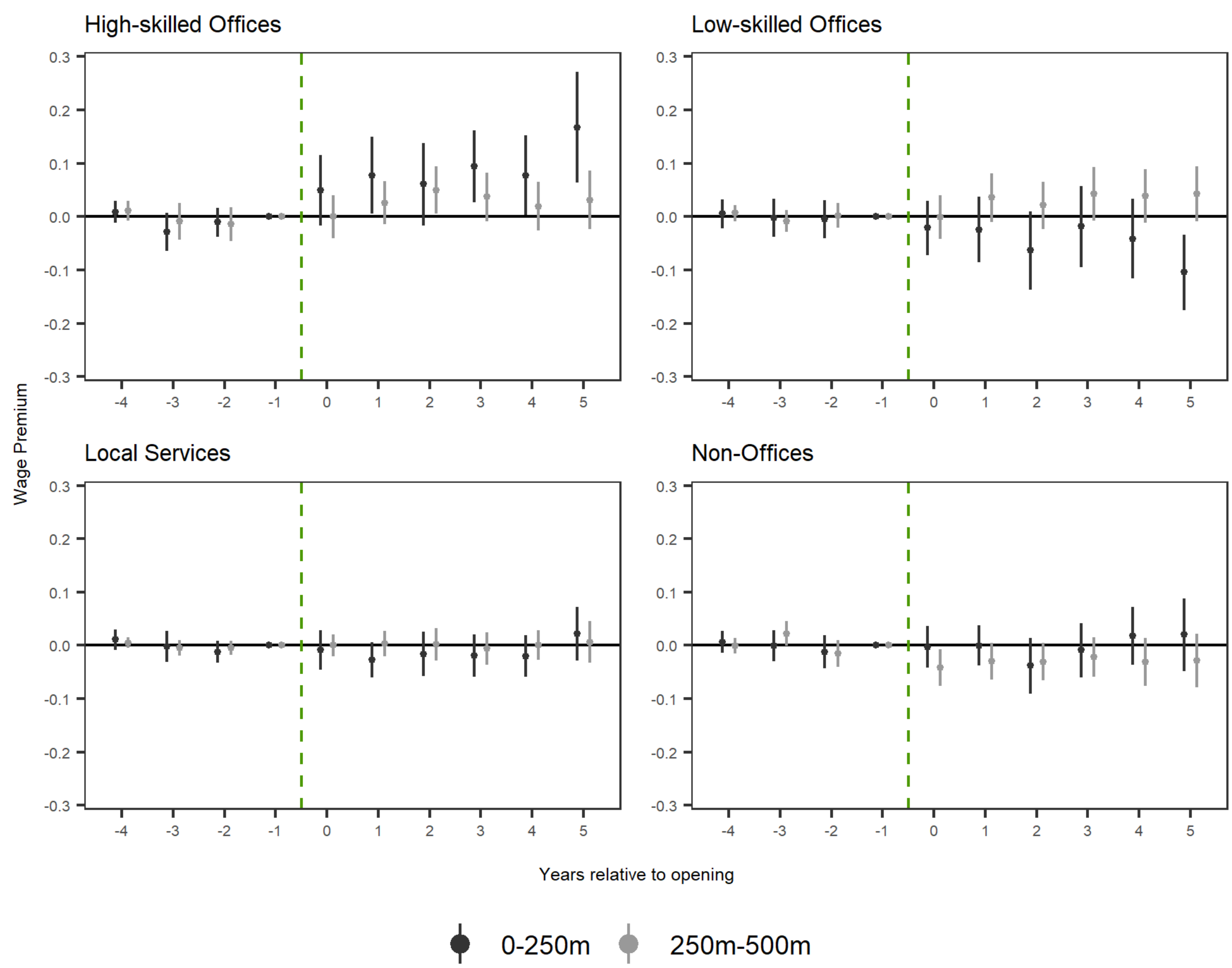


Note: This figure plots coefficients from running Equation (13) on log employment for different sectors. The definition of each sector is described in Section 2.2. The bars indicates the 95% confidence interval, where standard errors are clustered at the cell level.

workers in each sector are affected. Figure 8 indicates that the average wage premium increases differentially in first-ring cells for high-skilled offices but not for other sectors. There is a rise in the point estimates after the treatment starts that persists over the period of analysis. High-skilled offices also experience a differential increase in the share of college-educated workers in first-ring cells, as displayed in Figure 9. In this case, non-offices appear to be impacted as well.

Figures 8 and 9 reinforce the idea that new buildings impact the productivity of high-skilled offices. In particular, the model predicts that a local productivity shock would lead to higher wages in the affected sectors. On the other hand, local services would also observe higher wages due

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



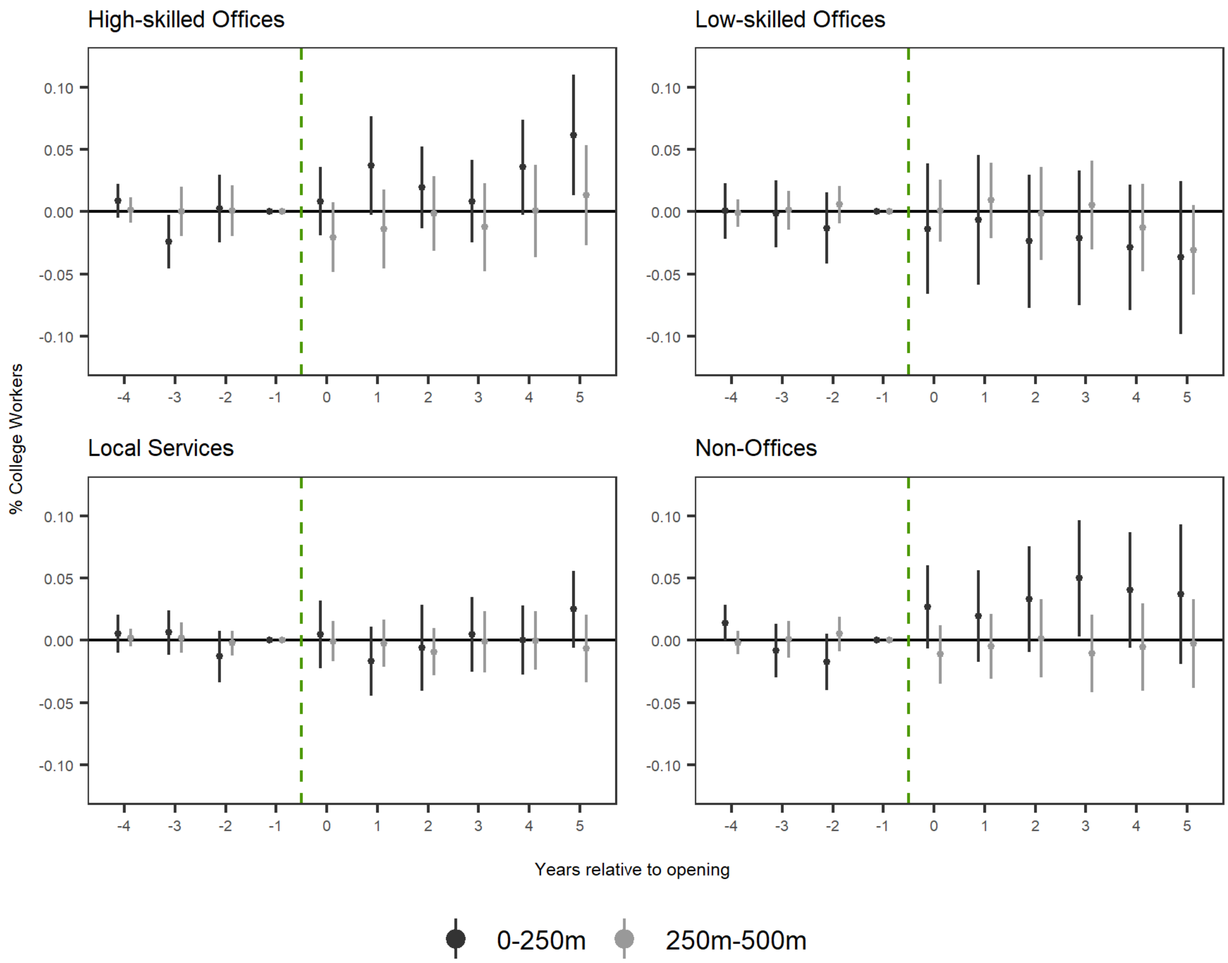
Note: This figure plots coefficients from running Equation (13) on mean establishment wage premium (weighted by establishment size) for different sectors. The definition of each sector is described in Section 2.2. The bars indicates the 95% confidence interval, where standard errors are clustered at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

to increased local demand, but Figure 8 shows no evidence of that. One possible explanation for this fact is the existence of heterogeneities in labor supply that are 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 turn to the results from Equation (14) for the outcomes considered thus far, summarized in Table 3. Panel A reports the effects on the aggregate economy. Consistent with the event-study

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



Note: This 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.

plots, Columns (1) and (2) exhibit impacts of 5% and 17.2% in the number of establishments and workers, respectively, for first-ring cells. Likewise, Columns (3) and (4) show that the share of college-educated workers and wages were unaffected.

Panels B and D exhibit the effects of new commercial buildings on high-skilled offices and local services, respectively. Regarding the first, there is an 16% increase in establishments and a 31.2% increase in employment in first-ring cells. High-skilled offices also experience an increase of 4 percentage points in the share of college-educated workers and an increase of 11.7% in the average wage premium. For local services, establishments and employment increase by 8.3% and

18.5%, respectively. It is worth noting that the higher coefficients for employment, compared to those for establishments, indicate that the differential growth of both sectors occurs at the intensive and extensive margins, that is, more establishments and more workers per establishment.

From the coefficients shown in Table 3, it is possible to derive a multiplier effect of the tradable sector on the non-tradable sector. From the baseline values, first-ring cells observe an increase of 47 workers in local services and 131 workers in high-skilled offices, meaning that roughly one job is created in local services for every three additional jobs in high-skilled offices. This number is 4.5 times lower than the one obtained by Moretti (2010), which estimates a local multiplier at the city level. 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 connections are symmetric, except when considering the direct effect of new buildings. 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).

I analyze this possibility in more detail in Supplementary Appendix F by estimating the local effects of new residential buildings. Since there is no reason to expect these constructions to affect local productivity, tradable sectors are unlikely to expand in treated neighborhoods. Any impact on local services, therefore, would be more plausibly interpreted as a direct effect of increased local demand. Table F6 shows weak evidence of greater economic activity in treated cells, and in particular, weak evidence of effects on local services. These findings support the interpretation that local services are indirectly affected by commercial buildings, primarily through an increase in high-skilled offices.[34]

[34]One might also be concerned about whether the results indeed stem from commercial buildings. A plausible alternative hypothesis is that new commercial buildings are correlated with surges in residential density. Consequently,

Table 3. Effects of New Commercial Buildings: DiD Results

	0-250m				250-500m			
	Log Estabs (1)	Log Workers (2)	% College (3)	Wage Premium (4)	Log Estabs (5)	Log Workers (6)	% College (7)	Wage Premium (8)
Panel A. All Sectors								
α_r	0.0498*	0.1715***	-0.0031	0.0047	0.0325*	0.0299	-0.0098	-0.0039
	(0.0262)	(0.0639)	(0.0155)	(0.0197)	(0.0185)	(0.0452)	(0.0084)	(0.0130)
R^2	0.01706	0.03050	-0.00035	-0.00027	0.00743	0.00066	0.00194	-0.00007
Obs	1,890	1,890	1,890	1,890	4,140	4,140	4,140	4,140
Panel B. High-Skilled Offices								
α_r	0.1598***	0.3123***	0.0403**	0.1167***	0.0409	0.1611*	0.0007	0.0286
	(0.0547)	(0.1134)	(0.0164)	(0.0394)	(0.0399)	(0.0910)	(0.0151)	(0.0204)
R^2	0.03736	0.03239	0.01903	0.04969	0.00199	0.00667	-0.00024	0.00339
Obs	1,890	1,890	1,890	1,890	4,140	4,140	4,140	4,140
Panel C. Low-Skilled Offices								
α_r	0.0161	0.1450	-0.0272	-0.0668**	0.0674*	0.0862	-0.0137	0.0342*
	(0.0521)	(0.1616)	(0.0240)	(0.0287)	(0.0372)	(0.0992)	(0.0142)	(0.0207)
R^2	-0.00004	0.00347	0.00537	0.02150	0.00780	0.00142	0.00139	0.00557
Obs	1,890	1,890	1,890	1,890	4,140	4,140	4,140	4,140
Panel D. Local Services								
α_r	0.0830**	0.1846***	0.0103	0.0003	0.0261	0.0312	-0.0044	0.0027
	(0.0332)	(0.0575)	(0.0129)	(0.0187)	(0.0219)	(0.0428)	(0.0093)	(0.0132)
R^2	0.02035	0.03724	0.00169	-0.00053	0.00235	0.00079	0.00017	-0.00016
Obs	1,890	1,890	1,890	1,890	4,140	4,140	4,140	4,140
Panel E. Non-Offices								
α_r	-0.0106	0.0906	0.0354*	0.0060	-0.0005	-0.0596	-0.0045	-0.0300*
	(0.0476)	(0.0777)	(0.0200)	(0.0241)	(0.0307)	(0.0616)	(0.0131)	(0.0175)
R^2	-0.00032	0.00370	0.01203	-0.00031	-0.00024	0.00127	-0.00004	0.00426
Obs	1,890	1,890	1,890	1,890	4,140	4,140	4,140	4,140

Notes: This table reports estimates of α_r in Equation (14) for different outcome variables and samples 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.

The second caveat is that São Paulo has a significant number of informal firms that are not included in the analysis. Given that informality is likely more prevalent among local services, the employment effects in this sector (and the associated multiplier effect) may be underestimated. However, because the new commercial buildings are located in the most prestigious areas of São Paulo, which are relatively more expensive, informality may play a negligible role in these neighborhoods.

offices and local services might locate nearby to benefit from proximity to a growing labor supply or greater local demand. However, Figure F2 suggests that the locations of new large commercial and residential buildings are weakly correlated.

5.3 Firm sorting Within Sectors

An important related issue is the extent to which changes in firm composition drive the findings of the previous section. In other words, does the rise in employment and wages result from the expansion of incumbent establishments, or is it driven by newly arrived ones? Furthermore, are these new establishments relocating from other neighborhoods, or are they being newly created? This section addresses these questions.

Leveraging the panel structure of the establishment-level data, I construct alternative aggregations of the outcome variables at the cell level based on two criteria: whether the establishment was already present in the cell before the opening of the closest new building, and for those that arrived afterward, their prior status before settling in the cell. I distinguish three cases: i) establishments relocating from treated or control cells, ii) establishments relocating from elsewhere in the São Paulo Metropolitan Area, and iii) newly created establishments.[35]

Table 4 presents estimates of Equation 14 using these alternative samples for high-skilled offices, which are the focus of my analysis. Results for other sectors in first-ring cells can be found in Tables E3, E4 and E5 in the Supplementary Appendix E. Panel A replicates the baseline results from the previous section, while the other panels present the new estimates. Unfortunately, because I exclude cells with at least one zero observation, the estimates are not fully comparable. However, they can still provide valuable information about the contribution of different groups of establishments to the observed effects on economic activity.

In Panel B, I present estimates considering only the establishments that existed before the opening of the closest new building. Column (1) shows a negative effect on the number of incumbent establishments, indicating increased establishment turnover in first-ring cells. Column (2) suggests that employment also declines, though the coefficient is not statistically significant at the 10% level. Columns (3) and (4) show no significant effects on worker composition or wages.

In Panel C, I expand the analysis to include newly created establishments. Columns (1) and

[35] It is possible that some establishments classified as new are actually relocations from other Brazilian cities. I do not differentiate between these cases.

Table 4. Effects on High-Skilled Offices: Firm Sorting and Relocation

	0-250m				250-500m			
	Log Estabs (1)	Log Workers (2)	% College (3)	Wage Premium (4)	Log Estabs (5)	Log Workers (6)	% College (7)	Wage Premium (8)
Panel A. All Establishments								
α_r	0.1598***	0.3123***	0.0403**	0.1167***	0.0409	0.1611*	0.0007	0.0286
	(0.0547)	(0.1134)	(0.0164)	(0.0394)	(0.0399)	(0.0910)	(0.0151)	(0.0204)
R^2	0.03736	0.03239	0.01903	0.04969	0.00199	0.00667	-0.00024	0.00339
Obs	1,890	1,890	1,890	1,890	4,140	4,140	4,140	4,140
Panel B. Only Incumbent Establishments								
α_r	-0.2421***	-0.2034	0.0060	0.0246	-0.1971***	-0.2990***	0.0086	0.0176
	(0.0646)	(0.1379)	(0.0240)	(0.0364)	(0.0502)	(0.0911)	(0.0181)	(0.0205)
R^2	0.07715	0.01197	-0.00026	0.00207	0.05004	0.02854	0.00032	0.00120
Obs	1,590	1,590	1,590	1,590	2,910	2,850	2,850	2,850
Panel C. Excluding Relocations								
α_r	-0.0129	0.0135	0.0256	0.0603*	-0.0277	-0.0337	0.0115	0.0107
	(0.0511)	(0.1192)	(0.0198)	(0.0349)	(0.0381)	(0.0852)	(0.0139)	(0.0192)
R^2	-0.00028	-0.00051	0.00680	0.01516	0.00087	0.00007	0.00089	0.00027
Obs	1,740	1,740	1,740	1,740	3,840	3,780	3,780	3,780
Panel D. Excluding Relocations within 1 km of New Buildings								
α_r	0.0830**	0.1846***	0.0103	0.0003	0.0261	0.0312	-0.0044	0.0027
	(0.0332)	(0.0575)	(0.0129)	(0.0187)	(0.0219)	(0.0428)	(0.0093)	(0.0132)
R^2	0.02035	0.03724	0.00169	-0.00053	0.00235	0.00079	0.00017	-0.00016
Observations	1,890	1,890	1,890	1,890	4,140	4,140	4,140	4,140

Notes: This table reports estimates of α_r in Equation (14) for different outcome variables and samples 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.

(2) show no significant effects on establishments or employment, and Columns (3) and (4) suggest some positive effects on worker composition and, more notably, on wages. Interestingly, there is a substantial increase in all point estimates compared to Panel B, implying that newly created establishments play an important role in explaining the baseline estimates.

Finally, panel D includes establishments that relocate from cells other than treated and control cells, that is, cells not displayed in panel A of Figure 4. Again, point estimates increase noticeably, particularly for establishments and employment, as shown in Columns (1) and (2). However, it is worth noting that these point estimates remain considerably smaller than those from the baseline estimates. This difference highlights the key role of relocations between treated and control cells in driving the results of Section 5.2.[36]

[36] These internal relocations can be seen as a typical violation of the Stable Unit Treatment Value Assumption (SUTVA). However, as discussed in Section 4.2, my estimates are designed to capture differential effects between neighborhoods with varying exposure to new commercial buildings. From both a conceptual and a policy perspective, there is no reason to consider firm relocation as an undesirable side effect.

Overall, Table 4 highlights the role of firm relocation in explaining spatial concentration within a city — a pattern also observed for local services, as shown in Table E4. Moreover, it reveals that the effects on wages and worker composition are driven primarily by post-new building establishments — those that are newly created or relocated. This finding, as discussed in Section 3.3, suggests that more productive high-skilled offices are more sensitive to productivity spillovers, thus enhancing the spatial sorting of firms.[37]

5.4 Floor Space Supply

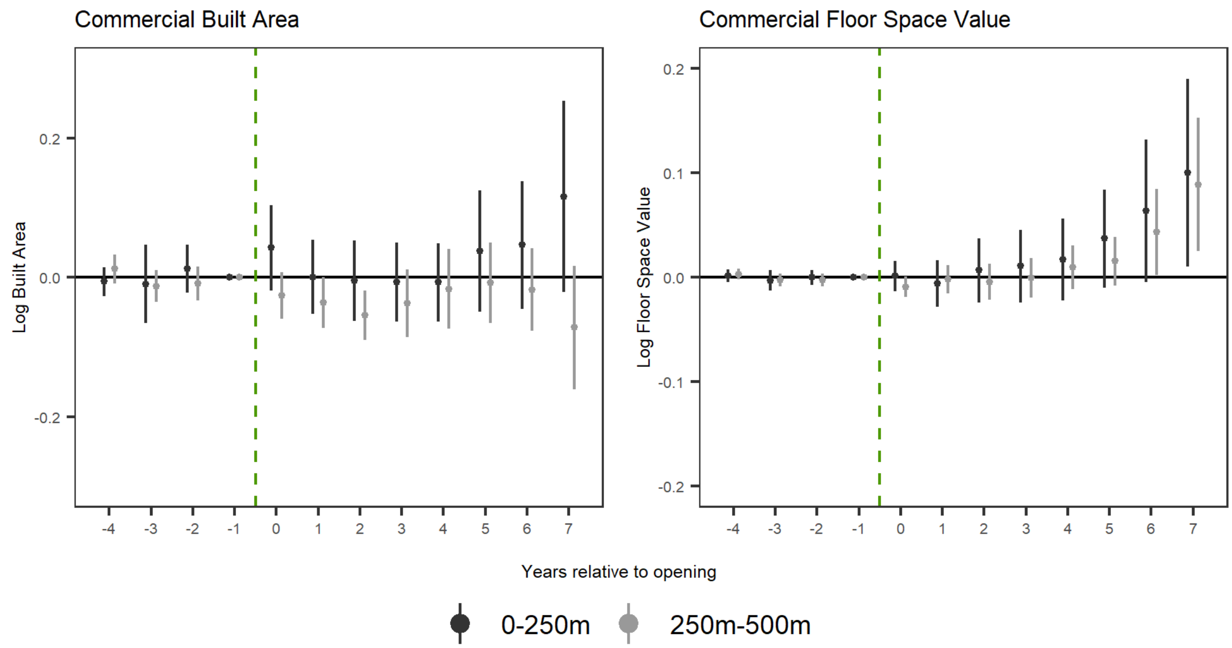
As the number of establishments differentially increases in first-ring cells, it is worth asking whether the floor space supply and prices respond to this local demand shift. To investigate this possibility, I estimate the effects of new buildings on the stock of commercial floor space and the average square meter value computed from IPTU data.

Figure 10 provides suggestive evidence of a differential increase in both variables in the medium run, five years after opening. This timing contrasts with the effects on economic activity, which materialize more quickly.[38]. However, as mentioned in footnote 24, there is a delay of about two years between a building's completion and its inclusion in the IPTU data, which probably exacerbates this disparity. Another important caveat concerns my measure of floor space value. While this variable reasonably correlates with rent prices in a cross-sectional spatial analysis, it is primarily used for tax purposes and does not fully reflect the actual market price. In particular, changes over time can be influenced by political factors, which could limit how accurately it reflects price dynamics.

[37]Some of the newly created establishments may in fact be informal firms transitioning into formality. Although I cannot disentangle this channel from true establishment creation, such transitions are rare (La Porta and Shleifer, 2014). Nevertheless, higher rates of formalization in more exposed cells can be interpreted as evidence of increased economic activity.

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

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



Note: This 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.

5.5 Discussion - Productivity Spillovers

The findings of this section show that urban concentration is driven both by firms that produce non-tradable goods and depend on local demand, and by firms that produce tradable goods. The increasing presence of the latter is less obvious and is consistent with the existence of local productivity spillovers. This interpretation becomes more plausible when 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 provides limited support for alternative explanations. For example, one might argue that neighborhoods near new buildings become more appealing to workers if these developments trigger improvements in local amenities, or if their location is closely tied to public infrastructure investments. In particular, new buildings could be a consequence of neighborhoods undergoing urban renewal. In such cases, more firms would choose to locate in these neighborhoods, as they would find it easier to recruit workers (Tsivanidis, 2023; Perez et al., 2022). High-skilled offices might benefit especially since they are more reliant on skilled workers, which are relatively scarce. However, under this scenario, the model implies a decrease in wages — a prediction that contrasts with the results. While I cannot entirely dismiss the relevance of this channel, the findings indicate that its role is likely to be secondary.[39]

Another possible explanation is that the sector classification proposed in this paper may be misleading and that the observed effects on high-skilled offices are instead driven by local demand linkages between firms. Yet, if this were the case, the effects on low-skilled offices should be just as large as those on local services and high-skilled offices, since they would likely depend on similar linkages.

However, it is conceivable that being in a high-employment neighborhood raises the likelihood of closing more deals for high-skilled offices in other ways. For example, if physical proximity attenuates information frictions between firms (Arzaghi and Henderson, 2008; Wu et al., 2020), or if being in an expensive location improves firm's reputation (Cook et al., 2007; Glückler, 2007), then location might be a crucial factor influencing demand in this sector. Nevertheless, while these mechanisms do not fit into the standard productivity spillover explanation, they are still based on externalities enhanced by density.

6 Industry Composition Within High-skilled Offices

I now examine how a neighborhood's industry composition interacts with that of nearby new buildings, focusing on the high-skilled office sector. This analysis is guided by two empirical questions: (i) to what extent the effects on more exposed neighborhoods are driven by industries that are more heavily represented in new buildings, and (ii) whether the industry composition of these neighborhoods shifts toward that of the new buildings after their openings. Importantly, I do

[39] Some papers investigate endogenous amenities effects using residential buildings as a shock. For example, Asquith et al. (2021) and Pennington (2021) find a decrease in rent prices, suggesting that this channel is of secondary importance. Diamond and McQuade (2019), on the other hand, show evidence of heterogeneous effects depending on the income level of the neighborhood. They report an increase in house prices in low-income neighborhoods and a decline in high-income neighborhoods.

not claim a causal relationship whereby the industry composition of new buildings influences that of nearby neighborhoods, as the former is undoubtedly endogenous. Instead, my goal is to assess whether there is a connection between the two that depends on physical proximity.

A key challenge in this inquiry is developing a generalizable measure that captures how a neighborhood's industry composition evolves relative to nearby new buildings. To address this, I propose the following relative composition index (RCI)[40]

$$RCI_{c,t} = Sh_{c,t}^{T} \cdot SimMatrix \cdot Sh_{b(c)} \quad . \tag{15}$$

In this expression, $RCI_{c,t}$ is the index for cell c in year t. $Sh_{c,t}^{T}$ is a vector of industry employment shares in cell c and year t, while $Sh_{b(c)}$ is a vector of industry employment shares for the associated building shock $b(c)$. $SimMatrix$ is a similarity matrix with measures of bilateral relatedness among industries.[41] Note that, while the cell employment shares are central to the analysis and computed for each period, I impose an invariant value for the building shock shares.

Equation (15) can be interpreted as a weighted sum of cell employment shares, with fixed weights given by the matrix product of $SimMatrix$ and $Sh_{b(c)}$. Let these weights be denoted by ω_i for industries $i \in I$. Thus, the RCI of a neighborhood increases when the employment of industries with higher ω_i grows relatively more, that is, industries that are more observed in the new buildings nearby. If there is an industry i' such that $\omega_{i'} > \omega_i$ for all $i \in I$, it follows that this expression is maximized when all workers in cell c are employed in industry i'.[42]

Industries are defined based on the 2-digit CNAE. To compute $Sh_{b(c)}$, I use the average

[40] A straightforward approach would be to focus on specific industries, such as finance or law. However, not only would this strategy reduce the scope of the analysis, but it would also limit the conclusions. Consider, for example, a neighborhood close to a new building occupied mostly by software engineering companies. If the finance sector grows relatively more in this neighborhood than in those further away, does this imply that industry composition does not play a role? Finance certainly shares more commonalities with software engineering than, say, law (such as labor skill requirements). Hence, narrowing the analysis may result in misleading conclusions.

[41] If there are I industries in this economy, $Sh_{c,t}^{T}$ is $(1 \times I)$, $SimMatrix$ is $(I \times I)$ and $Sh_{b(c)}$ is $(I \times 1)$.

[42] Consider the simplest case where $SimMatrix$ is a diagonal matrix, indicating that industries have no similarity to one another. In this case, the weights are simply given by the building's employment shares. If the largest employment share in a new building is, for example, software engineering, the RCI for a cell close to this building is maximized when software engineering is the sole industry present there.

employment shares in new buildings over the five years following their openings. If a cell is close to more than one building within the closest distance bin, I compute the employment shares using all of them. For the similarity matrix, I use two definitions: one based on worker flows, proposed by Jara-Figueroa et al. (2018), and another based on the correlation of occupation shares, proposed by Ellison et al. (2010). Supplementary Appendix D contains the details of how I construct these matrices.

6.1 Heterogeneous Effects Based on Industry Composition

To examine how the effects on high-skilled offices relate to industry composition, I propose the following modification of Equation (14):

$$y_{c,t} = \alpha_{r,high} D^{high}_{c,t,r} + \alpha_{r,low} D^{low}_{c,t,r} + \Psi_c + \mu_{d,t} + \upsilon_{c,t} \quad , \tag{16}$$

where $D^{high}_{c,t,r}$ and $D^{low}_{c,t,r}$ are indicators for treated cells whose RCI in the year before treatment is above or below the median index within the treated group, respectively.

Table 5 presents the results for the two proposed $SimMatrix$. Columns (1)-(4) show that first-ring cells with RCI above the median in worker flows experience greater effects on employment and the fraction of college-educated workers, smaller effects on the number of establishments, and no significant difference in wages. Second-ring cells above the median also exhibit larger employment effects. This pattern largely holds in Panel B, which considers occupational similarity, except that wage effects are now concentrated in cells below the median index. Overall, these findings suggest that neighborhoods with industries more prevalent in new buildings in the vicinity are slightly more affected.

6.2 Industry sorting

The other side of the coin is whether the building shock triggers changes in industry composition in more exposed cells. To do so, I now explore how the RCI varies over time, using the indices as

Table 5. Heterogeneous Effects on High-Skilled Offices based on Industry Composition

	0-250m				250-500m			
	Log Estabs (1)	Log Workers (2)	% College (3)	Wage Premium (4)	Log Estabs (5)	Log Workers (6)	% College (7)	Wage Premium (8)
Panel A. Worker Flows								
High RCI	0.1089	0.4384**	0.0488*	0.1140**	0.0448	0.2272**	-0.0264	0.0316
	(0.0742)	(0.1853)	(0.0258)	(0.0541)	(0.0472)	(0.0999)	(0.0188)	(0.0268)
Low RCI	0.2073***	0.1949	0.0325	0.1192**	0.0374	0.1015	0.0251	0.0259
	(0.0786)	(0.1329)	(0.0205)	(0.0570)	(0.0628)	(0.1472)	(0.0228)	(0.0302)
R^2	0.04258	0.03966	0.02019	0.04972	0.00201	0.00817	0.00762	0.00344
Obs	1,890	1,890	1,890	1,890	4,140	4,140	4,140	4,140
Panel C. % Occupations								
High RCI	0.1029	0.3833**	0.0418*	0.0584	0.0427	0.3306***	-0.0090	0.0555*
	(0.0733)	(0.1532)	(0.0251)	(0.0458)	(0.0502)	(0.1094)	(0.0180)	(0.0283)
Low RCI	0.2154***	0.2430	0.0389*	0.1736***	0.0394	0.0089	0.0094	0.0045
	(0.0790)	(0.1668)	(0.0211)	(0.0613)	(0.0609)	(0.1392)	(0.0236)	(0.0287)
R^2	0.04418	0.03481	0.01906	0.06748	0.00199	0.01652	0.00076	0.00748
Obs	1,890	1,890	1,890	1,890	4,140	4,140	4,140	4,140

Notes: This table reports estimates of $\alpha_{r,high}$ and $\alpha_{r,low}$ in Equation (16) for different outcome variables and samples 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.

outcome variables in expression (14).

The results in Table 6 indicate that the RCI based on worker flows decreases in first-ring cells, as shown in Column (1) and (2). Column (4) also shows a decline in the RCI based on occupational shares, although in this case the coefficient is not statistically significant at the 10% level. For second-ring cells, there is no meaningful evidence of changes in relative composition.

In summary, the findings in Tables 5 and 6 indicate that productivity spillovers are not limited to closely related firms. Neighborhoods are affected by new buildings in the vicinity regardless of their industry composition, which tend to be more diverse after openings. This evidence is consistent with earlier cross-city studies (Henderson et al., 1995; Duranton and Puga, 2001) and the more recent work on agglomeration patterns within cities by Baum-Snow et al. (2021). Importantly, my findings are not at odds with industry concentration forces, but rather indicate that local productivity spillovers may promote industry diversity.

Table 6. Industry Sorting within High-Skilled Offices

	0-250m		250-500m	
	RCI - Worker Flows	RCI - % Occupations	RCI - Worker Flows	RCI - % Occupations
	(1)	(2)	(3)	(4)
α_r	-0.0109**	-0.0074	0.0035	-0.0045
	(0.0052)	(0.0066)	(0.0048)	(0.0059)
R^2	0.01235	0.00293	0.00073	0.00076
Obs	1,890	1,890	4,140	4,140

Notes: This table reports estimates of α_r in Equation (14) for different outcome variables and samples 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.

7 Robustness Checks

This section summarizes a series of robustness tests to validate the findings reported in this paper. Supplementary Appendix F presents the results and provides additional methodological details. Some specific exercises indicate a smaller impact on local services employment — suggesting a potential overestimation of the multiplier effect — and a lower impact on high-skilled office wages. However, the overall results remain consistent.

Alternative Sample of Cells.—Focusing on neighborhoods with nonzero workers in all sectors and years, and at least one formal address, may be overly restrictive. Hence, I report additional estimates based on a larger sample of cells that do not impose these requirements. As shown in Table F1, the main results are preserved, although some differences emerge. First, local services now exhibit impacts on wages and worker composition. Second, there is a substantial increase in economic activity in both low-skilled office and non-office sectors.[43]

Alternative Sample of Buildings.—Because the selection of new buildings relies on *ad hoc* choices, I also check if 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-educated workers. Table F2 displays estimates of α_r from Equation (14) for different thresholds indicated in the columns, focusing exclusively on first-ring cells. The results remain largely unchanged, with an increase in economic activity driven

[43]Because some sector–cell–year combinations do not have economic activity, the estimates are obtained from an unbalanced panel.

by local services and high-skilled offices. High-skilled offices also experience increases in wages and the share of college-degree workers.

No Matching.—To understand to what extent the results are sensitive to the use of matching to construct the control groups, I also report estimates using the initial sample of 456 cells depicted in Panel A of Figure 4. In other words, I include all outer-ring cells as control units to estimate the effects on first- and second-ring cells. Table F3 presents a balance test for this sample. Compared to Table 2, the differences between the treated and control groups are much more pronounced. Outer-ring cells exhibit considerably less economic activity and their firms pay lower wages and are less intensive in skilled labor. Table F4 shows estimates of Equation (14) using this new sample. In general, the point estimates are lower than those in Table 3. Notably, the effects on the wages of high-skilled offices are of lower magnitude and statistical significance. The multiplier effect in this case is about 0.2, somewhat lower than what I obtain from Table 3.

Continuous Treatment Variable.—Tables F7 and F8 present results using a linear and an exponential continuous treatment variable, respectively. The conclusions remain unchanged.

Alternative Clustering of Standard Errors.—Table F9 displays estimates with standard errors clustered based on the nearest new commercial building, with very similar results.

Alternative Estimator.—I also report in Figures F3-F6 event-study plots using the doubly-robust estimator proposed by Callaway and Sant'Anna (2021). In this case, the figures exhibit qualitatively similar patterns observed in Figures 6-9, except that now the effects on establishments and employment for high-skilled offices appear to be concentrated on second-ring cells. Moreover, confidence intervals tend to be larger in general. I consider Gardner et al. (2024)'s estimator more suitable than Callaway and Sant'Anna (2021)'s in this setting for two reasons. First, in some years, the number of observations available to compute the group-time ATTs in Callaway and Sant'Anna (2021) is limited, leading to noisier estimates. Second, Gardner et al. (2024)'s approach provides greater flexibility in controlling for district-specific trends.

8 Conclusion

This paper explores the opening of large commercial buildings in São Paulo to study the concentration of economic activity within cities. I examine how different sectors are impacted in neighborhoods more exposed to these buildings. For this purpose, I develop a difference-in-differences identification strategy that combines the "ring" approach with a matching method to select treated and control neighborhoods.

The results indicate that neighborhoods within 250 meters of a new building experience a differential increase in employment, driven primarily by high-skilled offices and local services. I estimate that for every three additional jobs created by high-skilled offices, one job is created by local services. I also find evidence consistent with the hypothesis that the productivity of high-skilled offices is affected. There is a differential increase in wages and the share of college-educated workers within this sector. These effects seem to be largely driven by the sorting of high-wage firms into treated neighborhoods, either through the creation of new firms or the relocation of existing ones.

My findings support a description of spatial concentration based on productivity spillovers and local demand effects. Shifts in local productivity generated by spillovers increase incentives for offices to settle in affected neighborhoods. This expansion, in turn, leads to greater demand for local services and further increases incentives for the non-tradable sector to locate in these areas. In this sense, I interpret the estimated non-tradable/tradable growth ratio as a local multiplier effect.

Estimating a neighborhood local multiplier is informative for urban policies. For example, consider the persistent prevalence of remote and hybrid work due to the COVID-19 pandemic, especially among office workers (Barrero et al., 2023). As the number of commutes to the city center drops, the demand for local services in these neighborhoods decreases, and the local multiplier gives a notion of the magnitude of this shock.[44]

The results also highlight the importance of considering heterogeneities within and between

[44]On the other hand, residential areas may observe an increase in local services, since individuals now would spend more time in these locations. See Alipour et al. (2022).

sectors when modeling and estimating agglomeration forces. The literature on quantitative spatial models typically abstracts from these issues by assuming: i) common parameters across sectors; and ii) agglomeration effects that depend exclusively on local aggregate employment distribution. To what extent these approximations allow for accurate policy evaluations is an open question, and hopefully, this paper motivates more research on this front.

A Appendix

Proof of Proposition 1

To simplify the notation, I first rewrite expressions (7) and (9) as:

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

and

$$r_n = Q_n p_{LS,n}^{\kappa} \quad , \tag{A.2}$$

where $\kappa = \dfrac{(\chi-\delta\varepsilon)(1+\eta)}{1+\beta\chi(1+\eta)} > 0$. 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 $A_{s,n}$:

$$\xi \equiv \frac{A_{s,n}}{p_{LS,n}}\frac{\partial p_{LS,n}}{\partial A_{s,n}} = \frac{\nu_s A_{s,n}^{\chi(1+\eta)}}{\sum_{s'\neq LS}\nu_{s'}A_{s',n}^{\chi(1+\eta)}} \quad , \tag{A.3}$$

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 (4). Now, combining expressions (4) and (A.2) 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:

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

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

For local services, I first combine Equations (5) and (A.2). 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.

Proof of Proposition 2

To simplify the notation, I rewrite (3) for an office sector as

$$w_{s,n} = M p_s^{\frac{1}{1-\beta}} \left(\frac{A_{s.n}}{r_n^\beta} \right)^{\frac{1}{1-\beta}} \quad , \tag{A.4}$$

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 (A.4). 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 (A.3). Since $0 < \xi < 1$ and $0 < \beta\kappa < 1$, this derivative is positive.

For local services, the same procedure yields

$$w_{LS,n} = M \left(p_{LS,n}^{1-\beta\kappa} \right)^{\frac{1}{1-\beta}} \quad , \tag{A.5}$$

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

For real wages, I first use Equations (A.4) and (A.5) to get expressions for $w_{s,n}/p_{LS,n}^\delta$ and $w_{LS,n}/p_{LS,n}^\delta$:

$$w_{s,n}^R \equiv \frac{w_{s,n}}{p_{LS,n}^\delta} = M p_s^{\frac{1}{1-\beta}} \left(\frac{A_{s.n}}{p_{LS,n}^{\beta\kappa+\delta(1-\beta)}} \right)^{\frac{1}{1-\beta}} \quad , \tag{A.6}$$

and

$$w_{LS,n}^{R} \equiv \frac{w_{LS,n}}{p_{LS,n}^{\delta}} = M\left(p_{LS,n}^{1-\beta\kappa-\delta(1-\beta)}\right)^{\frac{1}{1-\beta}} \quad , \tag{A.7}$$

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

$$\begin{aligned}
\frac{\partial w_{s,n}^{R}}{\partial A_{s,n}} &= \frac{M}{1-\beta}p_{s}^{\frac{1}{1-\beta}}\left(\frac{A_{s.n}}{p_{LS,n}^{\beta\kappa+\delta(1-\beta)}}\right)^{\frac{1}{1-\beta}-1}\left[\frac{1}{p_{LS,n}^{\beta\kappa+\delta(1-\beta)}}-(\beta\kappa+\delta(1-\beta))\frac{A_{s.n}}{p_{LS,n}^{\beta\kappa+\delta(1-\beta)+1}}\frac{\partial p_{LS,n}}{\partial A_{s,n}}\right] \\
&= \frac{M}{1-\beta}p_{s}^{\frac{1}{1-\beta}}\left(\frac{A_{s.n}}{p_{LS,n}^{\beta\kappa+\delta(1-\beta)}}\right)^{\frac{1}{1-\beta}-1}\frac{1}{p_{LS,n}^{\beta\kappa+\delta(1-\beta)}}\left[1-(\beta\kappa+\delta(1-\beta))\xi\right] \quad ,
\end{aligned}$$

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 (A.7) shows that the derivative is positive if $\beta\kappa+\delta(1-\beta) < 1$, which is also not guaranteed.

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.

Alipour, J., Falck, O., Krause, S., Krolage, C., and Wichert, S. (2022). The future of work and consumption in cities after the pandemic: Evidence from germany. *CESifo Working Paper No. 10000*.

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*.

Atkin, D., Chen, K., and Popov, A. (2022). The returns to face-to-face interactions: Knowledge spillovers in silicon valley. *NBER Working Paper no. 30147*.

Barrero, J., Bloom, N., and Davis, S. (2023). The evolution of work from home. *NBER Working Paper no. 31686*.

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*.

Callaway, B. and Sant'Anna, P. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

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.

Cook, G., Pandit, N., Beaverstock, J., Taylor, P., and Pain, K. (2007). The role of location in knowledge creation and diffusion: evidence of centripetal and centrifugal forces in the City of London financial services agglomeration. *Environment and Planning A: Economy and Space*, 39(6):1325–1345.

Dauth, W., Findeisen, S. Moretti, E., and Suedekum, J. (2022). Matching in cities. *Journal of the European Economic Association*, 20(4):1478—-1521.

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.

Diamond, R. and McQuade, T. (2019). Who wants affordable housing in their backyard? An equilibrium analysis of low-income property development. *Journal of Political Economy*, 127(3):1063–1117.

Duranton, G. and Puga, D. (2001). Nursery cities: Urban diversity, process innovation, and the life cycle of products. *American Economic Review*, 91(5):1454–1477.

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

Ellison, G., Glaeser, E., and Kerr, W. (2010). What causes industry agglomeration? Evidence from coagglomeration patterns. *American Economic Review*, 100(3):1195—-1213.

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

Gardner, J., Thakral, N., Tô, L., and Yap, L. (2024). Two-stage differences in differences. *Working Paper*.

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

Glückler, J. (2007). Geography of reputation: The city as the locus of business opportunity. *Regional Studies*, 41(7):949–961.

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.

Guimaraes, P., Figueiredo, O., and Woodward, D. (2011). Accounting for neighboring effects in measures of spatial concentration. *Journal of Regional Science*, 51(4):678–693.

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.

Henderson, V., Kuncoro, A., and Turner, M. (1995). Industrial development in cities. *Journal of Political Economy*, 103(5):1067–1090.

Jaffe, A., Trajtenberg, M., and Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics*, 108(3):577—-598.

Jara-Figueroa, C., Jun, B., Glaeser, E., and Hidalgo, C. (2018). The role of industry-specific, occupation-specific, and location-specific knowledge in the growth and survival of new firms. *Proceedings of the National Academy of Sciences*, 115(50):12646–12653.

La Porta, R. and Shleifer, A. (2014). Informality and development. *Journal of Economic Perspectives*, 28(3):109–126.

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.

OECD/European Commission (2020). Cities in the world: A new perspective on urbanisation. Technical report, OECD Urban Studies, OECD Publishing, Paris.

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

Perez, J., Vial, F., and Zarate, R. (2022). Urban transit infrastructure: Spatial mismatch and labor market power. *CAF Development Bank Of Latinamerica - Research Department working papers 1992*.

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. (2023). Evaluating the impact of urban transit infrastructure: Evidence from Bogotá's Transmilenio. *Working Paper*.

Tsivanidis, N. and Gechter, M. (2023). Spatial spillovers from high-rise developments: Evidence from the Mumbai mills. *Working Paper*.

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

Wu, Y., Lee, C., Lee, C., and Peng, D. (2020). Geographic proximity and corporate investment efficiency: Evidence from high-speed rail construction in China. *Journal of Banking and Finance*, 140:106510.

Supplementary Appendix

Contents

B Multi-establishment Classification

For firms with at least 20 establishments located in São 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.

C 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 is located 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.[45]

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

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

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 $\phi_{m,e} = (B_{m,n(e)} w_e / p^{\delta}_{LS,n(e)})^{\eta}$. 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

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

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

[45] 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

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

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

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

Finally, I assume that the number of firms high enough so ϕ_m can be treated as fixed, and denote $B_{n(e)} \equiv \sum_m \frac{\overline{L}_m}{\phi_m} B^{\eta}_{m,n(e)}$ the Firm Commuter Market Access of neighborhood n. Equation (2) is then obtained.

Firm Location Choice.— To derive Equations (4) and (5), it is necessary first to compute the respective profit functions. I do so by using the first-order conditions of (1), which yields

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

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 (C.2). Finally, the number of firms from sector s that choose to locate in n is the product between this probability and $\overline{E}_s$

Market Clearing Conditions.— From the first order conditions of (1), 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 $f_{LS,n}$:

$$w_{s,n}\ell_{s,n} = \left[\frac{\varepsilon}{\chi\beta} \right]^{\varepsilon+1} \beta^{\chi} \frac{B_n}{p_{LS,n}^{\delta\varepsilon}} \left(\frac{p_s A_{s,n}}{r_n^{\beta}} \right)^{\chi} \quad , \tag{C.6}$$

$$w_{LS,n}\ell_{LS,n} = \left[\frac{\varepsilon}{\chi\beta} \right]^{\varepsilon+1} \beta^{\chi} \frac{B_n p_{LS,n}^{\chi-\delta\varepsilon}}{r_n^{\chi\beta}} \quad , \tag{C.7}$$

$$Y_{LS,n} = \frac{w_{LS,n}\ell_{LS,n}}{p_{LS,n}} \frac{\chi}{\varepsilon} \quad \text{and} \tag{C.8}$$

$$f_{s,n} = \left[\frac{\varepsilon}{\chi\beta} \right]^{\varepsilon} \beta^{\chi} \frac{B_n}{r_n^{\chi-\varepsilon} p_{LS,n}^{\delta\varepsilon}} (p_s A_{s,n})^{\chi} \quad . \tag{C.9}$$

$$f_{LS,n} = \left[\frac{\varepsilon}{\chi\beta} \right]^{\varepsilon} \beta^{\chi} \frac{B_n p_{LS,n}^{\chi-\delta\varepsilon}}{r_n^{\chi-\varepsilon}} \quad . \tag{C.10}$$

Combining (C.8) and (6) 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 (4), (5), (C.6) and (C.7) and solving for $p_{LS,n}$

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

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

To derive Equation (9), I first plug expressions (4), (5), (C.9) and (C.10) into (8):

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

Now, note that Equation (7) can be used to substitute the summation term inside the brackets:

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

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

Finally, I rearrange the terms to get

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

D Construction of Similarity Matrices

This section details about the similarity matrices used in the empirical analysis. These matrices are constructed from the same dataset used to estimate wage premiums but are restricted to establishments classified as high-skilled offices (see Section 2.1).

The similarity matrix based on worker flows is built on Jara-Figueroa et al. (2018). For pairs of industries i and i', I run the following regression:

$$F_{i \leftrightarrow i'} = \beta_0 + \beta_1 L_{i,i'} + \upsilon_{i,i'} \quad , \tag{D.11}$$

where $F_{i \leftrightarrow i'}$ represents the log of one plus the flow of workers between industries i and i', accounting for movement in both directions (making SimMatrix symmetric). $L_{i,i'}$ denotes the log of the total number of employment records in i and i', while $\upsilon_{i,i'}$ is the residual, normalized according to the following expression:

$$\omega^{WF}_{i,i'} = \begin{cases} \dfrac{\upsilon_{i,i'} - min\{\upsilon_{i,i'}\}}{max\{\upsilon_{i,i'}\} - min\{\upsilon_{i,i'}\}} & , \ i \neq i' \\ 1 & , \ i = i' \end{cases} \tag{D.12}$$

The similarity matrix is then composed of the set of $\tilde{\upsilon}_{i,i'}$ obtained through this procedure.

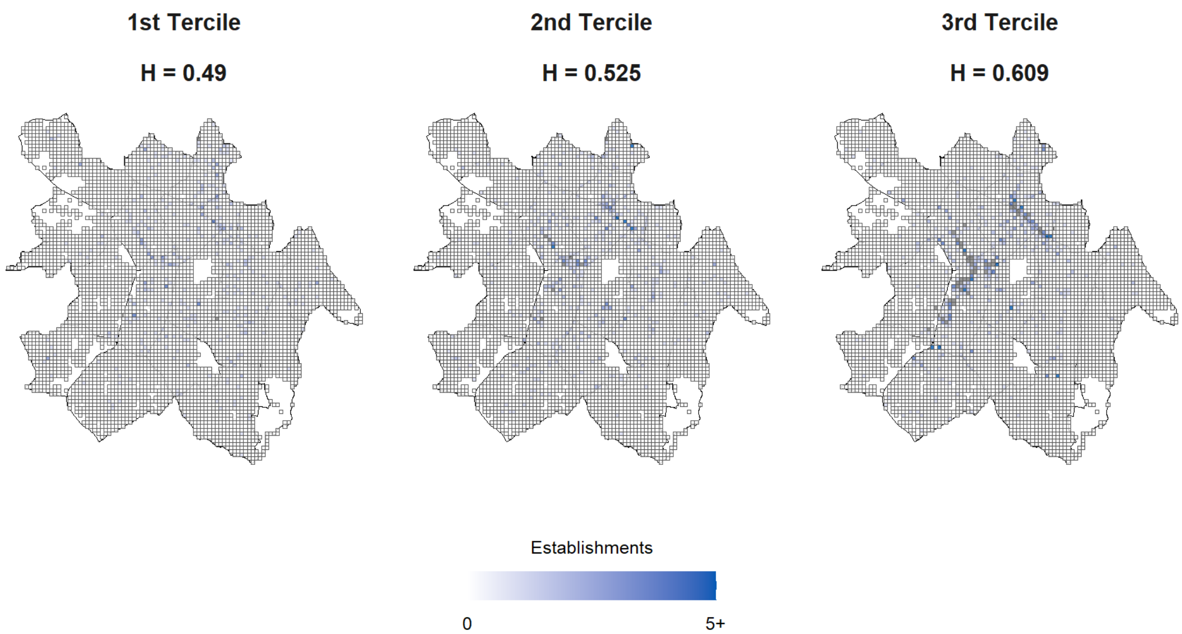
The similarity matrix based on occupational similarity follows Ellison et al. (2010) and uses 4-digit occupation data. For a given industry pair i-i', the similarity term $\omega^{OC}_{i,i'}$ is

$$\omega^{OC}_{i,i'} = max[corr(OcSh_i, OcSh_{i'}), 0] \quad , \tag{D.13}$$

where $corr(OcSh_i, OcSh_{i'})$ represents the correlation between the occupational share vectors of industries i and i'.

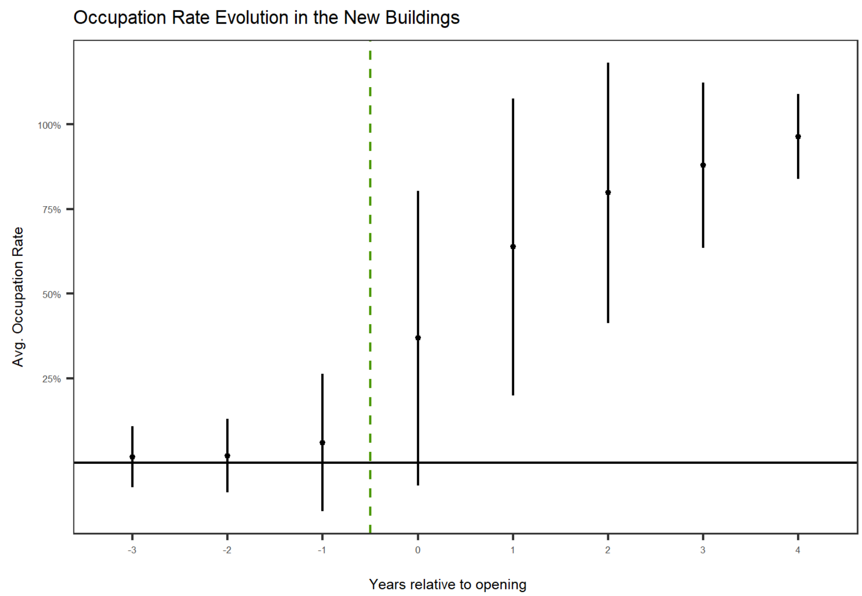
E Additional Figures and Tables

Figure E1. Spatial Distribution of Finance Establishments



Notes: This 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 E2. Evolution of Occupation Rate - New Buildings



Notes: This 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.

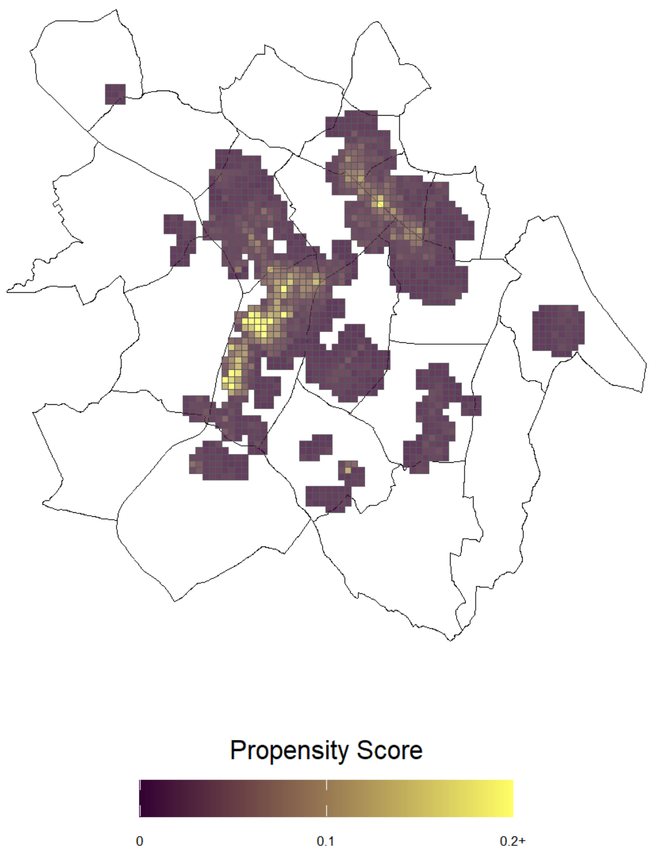
Figure E3. Examples of Commercial Buildings

Source: Google Maps.

Figure E4. Validating Selected Buildings

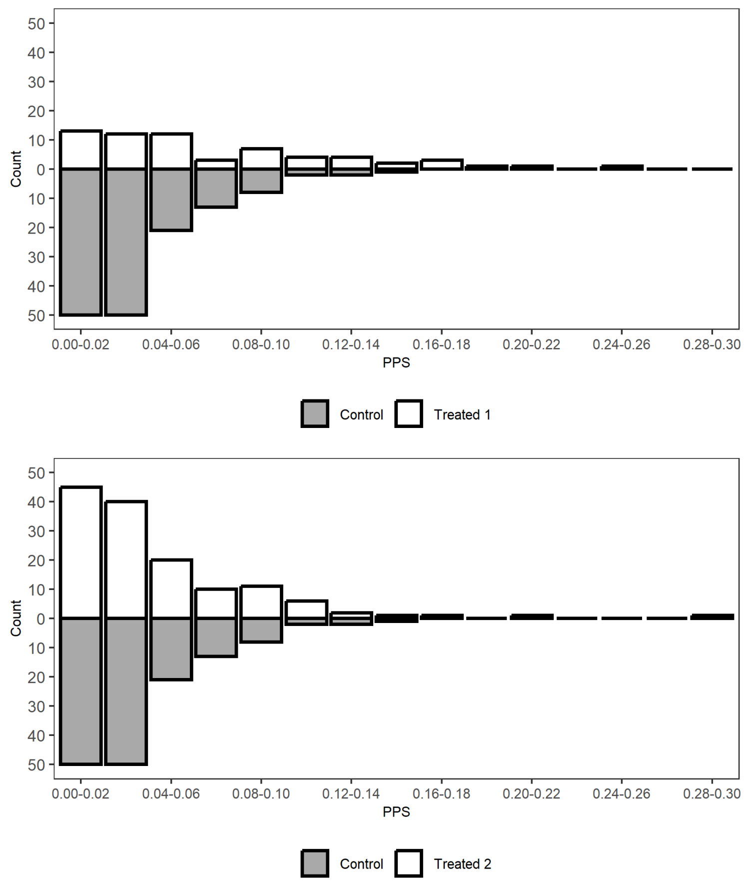
Source: Google Maps. *Notes:* Example of a commercial building inaugurated in 2013, located at Jornalista Roberto Marinho Avenue, 85, 04576-010. Using Google Maps imagery from 2011 (left) and 2014 (right), it is possible to check if the timeline of construction is consistent with the inauguration year defined in Section 4.

Figure E5. Fitted values: Propensity Score



Note: This figure shows the results of the Lasso estimation of the propensity score model.

Figure E6. Fitted values: Propensity Score



Note: This figure shows histograms of the Proximity Probability Score for treated and control groups.

Table E1. List of Variables Used in the Propensity Score Model

Variable	**Description**	**Data Source**
Employment - cell	log of (1 + employment) within the cell	RAIS 2005
Employment - buffer	log of (1 + employment) within 500 meters of the cell's centroid, excluding the cell itself	RAIS 2005
Industry employment - cell	log of (1 + employment) by industry within 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
Industry employment - buffer	log of (1 + employment) by sector within 500 meters of the cell's centroid, excluding the cell itself. For the list of industries, see the third row	RAIS 2005
Wage - buffer	log of mean wages of workers within 500 m of the cell's centroid	RAIS 2005
Employment growth - cell	Difference in the log of (1 + employment) between 2005 and 2003 within the cell	RAIS 2003-2005
Employment growth - buffer	Difference in the log of (1 + employment) between 2005 and 2003 within 500 meters of the cell's centroid, excluding the cell itself	RAIS 2003-2005
Industry employment growth - cell	Difference in the log of o(1 + employment) by industry between 2005 and 2003 within the cell. For the list of industries, see the third row	RAIS 2003-2005
Industry employment growth - buffer	Difference in the log of (1 + employment) by industry between 2005 and 2003 within 500 meters of the cell's centroid, excluding the cell itself. For the list of industries, see the third row	RAIS 2003-2005
Wage growth - buffer	Difference in the log of mean wages of workers within 500 meters of the cell's centroid between 2005 and 2003	RAIS 2003-2005
Commercial built area - cell	log of (1 + commercial floor space stock) within the cell	IPTU 2005
Commercial built area - buffer	log of (1 + commercial floor space stock) within 500 meters of the cell's centroid, excluding the cell itself	IPTU 2005
Residential built area - cell	log of (1 + residential floor space stock) within the cell	IPTU 2005
Residential built area - buffer	log of (1 + residential floor space stock) within 500 meters of the cell's centroid, excluding the cell itself	IPTU 2005
Commercial land area - cell	log of (1 + commercial land area) within the cell	IPTU 2005
Commercial land area - buffer	log of (1 + commercial land area) within 500 meters of the cell's centroid, excluding the cell itself	IPTU 2005
Residential land area - cell	log of (1 + residential land area) within the cell	IPTU 2005
Residential land area - buffer	log of (1 + residential land area) within 500 meters of the cell's centroid, excluding the cell itself	IPTU 2005
Vacant land area - cell	log of (1 + vacant land area) within the cell	IPTU 2005
Vacant land area - buffer	log of (1 + vacant land area) within 500 meters of the cell's centroid, excluding the cell itself	IPTU 2005
Commercial built-area-ratio (BAR) - buffer	Ratio of total commercial floor space to total commercial land area within 500 meters of the cell's centroid	IPTU 2005
Residential built-area-ratio (BAR) - buffer	Ratio of total residential floor space to total residential land area within 500 meters of the cell's centroid	IPTU 2005
Commercial built value - buffer	log of commercial floor space value per square meter within 500 meters of the cell's centroid	IPTU 2005
Residential built value - buffer	log of residential floor space value per square meter within 500 meters of the cell's centroid	IPTU 2005
Commercial land value - buffer	log of commercial land area value per square meter within 500 meters of the cell's centroid	IPTU 2005
Residential land value - buffer	log of residential land area value per square meter within 500 meters of the cell's centroid	IPTU 2005
Commercial built area growth - cell	Difference in the log of (1 + commercial floor space stock) between 2005 and 2003 within the cell	IPTU 2003-2005
Commercial built area growth - buffer	Difference in the log of (1 + commercial floor space stock) between 2005 and 2003 within 500 meters of the cell's centroid, excluding the cell itself	IPTU 2003-2005
Residential built area growth - cell	Difference in the log of (1 + residential floor space stock) between 2005 and 2003 within the cell	IPTU 2003-2005
Residential built area growth - buffer	Difference in the log of (1 + residential floor space stock) between 2005 and 2003 within 500 meters of the cell's centroid, excluding the cell itself	IPTU 2003-2005

Variable	**Description**	**Data Source**
Commercial land area growth - cell	Difference in the log of (1 + commercial land area) between 2005 and 2003 within the cell	IPTU 2003-2005
Commercial land area growth - buffer	Difference in the log of (1 + commercial land area) between 2005 and 2003 within 500 meters of the cell's centroid, excluding the cell itself	IPTU 2003-2005
Residential land area growth - cell	Difference in the log of (1 + residential land area) between 2005 and 2003 within the cell	IPTU 2003-2005
Residential land area growth - buffer	Difference in the log of (1 + residential land area) between 2005 and 2003 within 500 meters of the cell's centroid, excluding the cell itself	IPTU 2003-2005
Vacant land area growth - cell	Difference in the log of (1 + vacant land area) between 2005 and 2003 within the cell	IPTU 2003-2005
Vacant land area growth - buffer	Difference in the log of (1 + vacant land area) between 2005 and 2003 within 500 meters of the cell's centroid, excluding the cell itself	IPTU 2003-2005
Commercial built-area-ratio (BAR) growth - buffer	Difference in the commercial BAR within 500 meters of the cell's centroid between 2005 and 2003	IPTU 2003-2005
Residential built-area-ratio (BAR) growth - buffer	Difference in the residential BAR within 500 meters of the cell's centroid between 2005 and 2003	IPTU 2003-2005
Commercial built value growth - buffer	Difference in the log of commercial floor space value within 500 meters of the cell's centroid between 2005 and 2003	IPTU 2003-2005
Residential built value growth - buffer	Difference in the log of residential floor space value within 500 meters of the cell's centroid between 2005 and 2003	IPTU 2003-2005
Commercial land value growth - buffer	Difference in the log of commercial land area value within 500 meters of the cell's centroid between 2005 and 2003	IPTU 2003-2005
Residential land value growth - buffer	Difference in the log of residential land area value within 500 meters of the cell's centroid between 2005 and 2003	IPTU 2003-2005
Number of train and subway stations - buffer	Number of train and subway stations within 500 meters of the cell's centroid	SP Metro and CPTM
Population - buffer	Log of residents within 500 meters of the cell's centroid	2000 Census - tract level
Log households - buffer	Log of households within 500 meters of the cell's centroid	2000 Census - tract level
Log per capita income - buffer	Log of per capita income within 500 meters of the cell's centroid	2000 Census - tract level
% population 18-40 - buffer	Share of population between 18 and 40 years old within 500 meters of the cell's centroid	2000 Census - tract level
% population 41-60 - buffer	Share of population between 41 and 60 years old within 500 meters of the cell's centroid	2000 Census - tract level
% population non-white - buffer	Share of brown and black individuals within 500 meters of the cell's centroid	2000 Census - tract level
% renters - buffer	Share of rented households within 500 meters of the 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 within 500 meters of the 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 of a monthly minimum wage and less than one monthly minimum wage within 500 meters of the 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 times the monthly minimum wages within 500 meters of the cell's centroid	2000 Census - tract level
% per capita income > 3 min. wages - buffer	Share of households whose per capita income is greater than three times the monthly minimum wages within 500 meters of the cell's centroid.	2000 Census - tract level
Distance to the CBD	Log of the distance between the cell's centroid and Se Square (in km).	-
Employment to population ratio - buffer	Ratio of employment to resident population within 500 meters of the cell's centroid.	RAIS 2005 and Census 2000 - tract level

Table E2. Coefficients: Propensity Score Model

Variable	Estimate
Commercial built-area-ratio (BAR) – 500m buffer	0.542
Residential BAR growth– 500m buffer	−2.085
Log commercial floor space growth – cell	0.051
Log commercial floor space value growth – 500m buffer	10.447
Log (1 + employment adm. and support) – cell	0.099
Log (1 + employment construction) – cell	0.007
Log (1 + employment information and communication) – 500m buffer	0.020
Log (1 + employment information and communication) – cell	0.068
Log (1 + employment wholesale) – cell	0.055
Log residential land area growth – cell	0.188
Log (1 + vacant land area) – 500m buffer	0.037

Notes: This table shows the estimated coefficients of Equation (12) using Lasso. For a description of all variables used, see Table E1

Table E3. Effects on Low-Skilled Offices: Firm Sorting and Relocation

	0-250m				250-500m			
	Log Estabs (1)	Log Workers (2)	% College (3)	Wage Premium (4)	Log Estabs (5)	Log Workers (6)	% College (7)	Wage Premium (8)
Panel A. All Establishments								
α_r	0.0161	0.1450	-0.0272	-0.0668**	0.0674*	0.0862	-0.0137	0.0342*
	(0.0521)	(0.1616)	(0.0240)	(0.0287)	(0.0372)	(0.0992)	(0.0142)	(0.0207)
R^2	-0.00004	0.00347	0.00537	0.02150	0.00780	0.00142	0.00139	0.00557
Obs	1,890	1,890	1,890	1,890	4,140	4,140	4,140	4,140
Panel B. Only Incumbent Establishments								
α_r	-0.1836***	-0.1339	-0.0202	-0.0327	-0.1641***	-0.1714	-0.0011	0.0163
	(0.0646)	(0.1640)	(0.0281)	(0.0318)	(0.0438)	(0.1162)	(0.0176)	(0.0192)
R^2	0.05742	0.00337	0.00225	0.00532	0.04312	0.00687	-0.00029	0.00137
Obs	1,530	1,440	1,440	1,440	3,450	3,330	3,330	3,330
Panel C. Excluding Relocations								
α_r	-0.0756	-0.1062	-0.0325	-0.0822***	0.0330	0.0264	-0.0048	0.0390*
	(0.0519)	(0.1273)	(0.0240)	(0.0308)	(0.0348)	(0.0974)	(0.0146)	(0.0206)
R^2	0.00970	0.00203	0.00770	0.03190	0.00174	-0.00008	-0.00004	0.00751
Obs	1,860	1,830	1,830	1,830	4,080	4,020	4,020	4,020
Panel D. Excluding Relocations within 1 km of New Buildings								
α_r	-0.0445	0.0494	-0.0253	-0.0580*	0.0478	0.0650	-0.0082	0.0361*
	(0.0524)	(0.1584)	(0.0240)	(0.0309)	(0.0375)	(0.1056)	(0.0145)	(0.0202)
R^2	0.00302	-0.00006	0.00445	0.01514	0.00373	0.00068	0.00036	0.00650
Obs	1,890	1,890	1,890	1,890	4,110	4,050	4,050	4,050

Notes: This table reports estimates of α_r in Equation (14) for different outcome variables and samples 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.

Table E4. Effects on Local Services: Firm Sorting and Relocation

	0-250m				250-500m			
	Log Estabs (1)	Log Workers (2)	% College (3)	Wage Premium (4)	Log Estabs (5)	Log Workers (6)	% College (7)	Wage Premium (8)
Panel A. All Establishments								
α_r	0.0830**	0.1846***	0.0103	0.0003	0.0261	0.0312	-0.0044	0.0027
	(0.0332)	(0.0575)	(0.0129)	(0.0187)	(0.0219)	(0.0428)	(0.0093)	(0.0132)
R^2	0.02035	0.03724	0.00169	-0.00053	0.00235	0.00079	0.00017	-0.00016
Obs	1,890	1,890	1,890	1,890	4,140	4,140	4,140	4,140
Panel B. Only Incumbent Establishments								
α_r	-0.2133***	-0.1552**	-0.0078	-0.0195	-0.1653***	-0.1662***	-0.0118	0.0034
	(0.0491)	(0.0647)	(0.0162)	(0.0233)	(0.0357)	(0.0527)	(0.0100)	(0.0131)
R^2	0.09398	0.02431	0.00052	0.00301	0.05903	0.02591	0.00293	-0.00012
Obs	1,800	1,770	1,770	1,770	3,960	3,840	3,840	3,840
Panel C. Excluding Relocations								
α_r	0.0456	0.0895	-0.0010	-0.0107	0.0010	-0.0065	-0.0072	0.0001
	(0.0346)	(0.0587)	(0.0125)	(0.0193)	(0.0236)	(0.0421)	(0.0092)	(0.0131)
R^2	0.00575	0.00824	-0.00051	0.00070	-0.00024	-0.00020	0.00086	-0.00024
Obs	1,890	1,890	1,890	1,890	4,140	4,140	4,140	4,140
Panel D. Excluding Relocations within 1 km of New Buildings								
α_r	0.0646*	0.1332**	0.0043	-0.0050	0.0059	-0.0006	-0.0086	-0.0002
	(0.0349)	(0.0607)	(0.0128)	(0.0192)	(0.0228)	(0.0426)	(0.0091)	(0.0132)
R^2	0.01186	0.01831	-0.00016	-0.00027	-0.00011	-0.00024	0.00131	-0.00024
Obs	1,890	1,890	1,890	1,890	4,140	4,140	4,140	4,140

Notes: This table reports estimates of α_r in Equation (14) for different outcome variables and samples 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.

Table E5. Effects on Non-offices: Firm Sorting and Relocation

	0-250m				250-500m			
	Log Estabs (1)	Log Workers (2)	% College (3)	Wage Premium (4)	Log Estabs (5)	Log Workers (6)	% College (7)	Wage Premium (8)
Panel A. All Establishments								
α_r	-0.0106	0.0906	0.0354*	0.0060	-0.0005	-0.0596	-0.0045	-0.0300*
	(0.0476)	(0.0777)	(0.0200)	(0.0241)	(0.0307)	(0.0616)	(0.0131)	(0.0175)
R^2	-0.00032	0.00370	0.01203	-0.00031	-0.00024	0.00127	-0.00004	0.00426
Obs	1,890	1,890	1,890	1,890	4,140	4,140	4,140	4,140
Panel B. Only Incumbent Establishments								
α_r	-0.1604**	0.1244	0.0626**	0.0247	-0.1485***	-0.1584*	0.0165	-0.0583**
	(0.0778)	(0.1371)	(0.0265)	(0.0305)	(0.0402)	(0.0902)	(0.0165)	(0.0233)
R^2	0.03220	0.00458	0.03334	0.00255	0.03248	0.00775	0.00217	0.01468
Obs	1,530	1,470	1,470	1,470	3,570	3,420	3,420	3,420
Panel C. Excluding Relocations								
α_r	-0.1187**	-0.1005	0.0334	-0.0093	-0.0798**	-0.2307***	0.0043	-0.0379**
	(0.0473)	(0.0951)	(0.0213)	(0.0261)	(0.0321)	(0.0680)	(0.0137)	(0.0180)
R^2	0.02656	0.00398	0.00995	-0.00005	0.01122	0.02088	-0.00008	0.00672
Obs	1,860	1,860	1,860	1,860	4,080	4,020	4,020	4,020
Panel D. Excluding Relocations within 1 km of New Buildings								
α_r	-0.0352	0.0332	0.0332*	-0.0004	-0.0373	-0.1709**	-0.0085	-0.0369**
	(0.0456)	(0.0882)	(0.0201)	(0.0227)	(0.0320)	(0.0667)	(0.0130)	(0.0179)
R^2	0.00193	-0.00002	0.01083	-0.00054	0.00231	0.01158	0.00045	0.00661
Obs	1,860	1,860	1,860	1,860	4,140	4,110	4,110	4,110

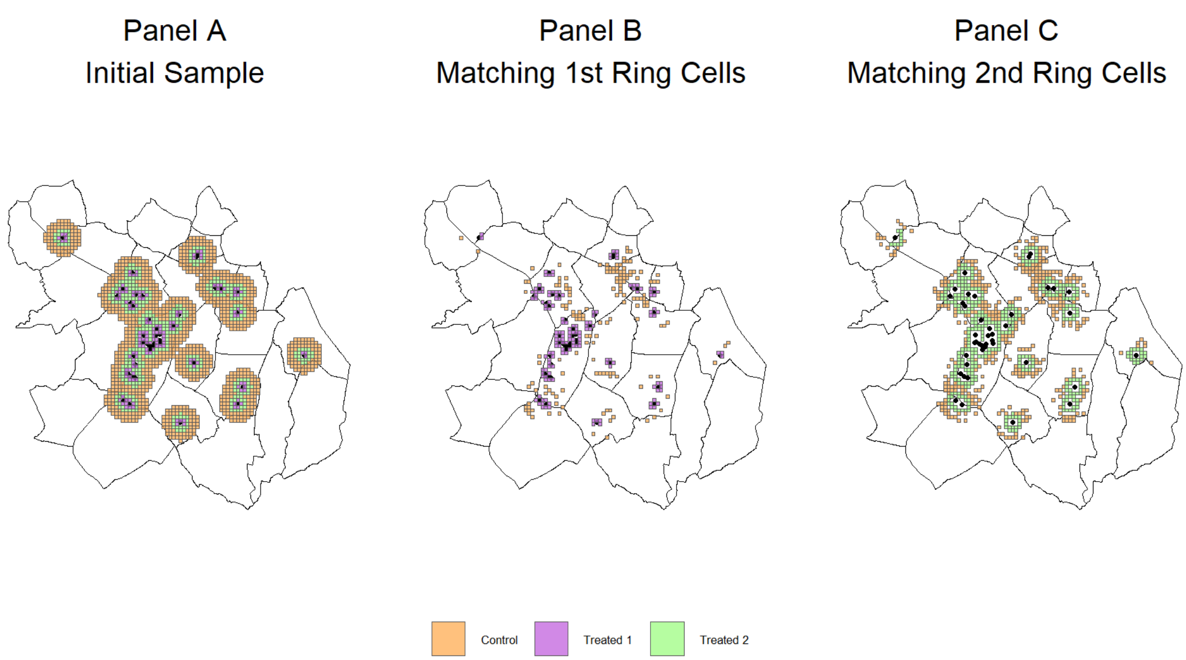
Notes: This table reports estimates of α_r in Equation (14) for different outcome variables and samples indicated in the columns. Standard errors clustered at the cell level are displayed in parentheses. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

F Robustness Checks

Larger Sample of Neighborhoods

Using now all cells within 1 km of a new building site, I identify treated and control groups and perform the matching in the same way described in Section 4. Figure F1 illustrates this procedure, and Table F1 presents the results from Equation (14), which replicates the structure of Table 3. Note that in this case, the samples used for each sector differ because some cells may not contain all sectors. Moreover, since zero or missing observations are dropped, the estimates are obtained from an unbalanced panel.

Figure F1. Empirical Analysis Setup



Notes: This figure depicts the design of the empirical analysis using an alternative sample of cells, as described in Section F. Panel A shows the initial sample of treated and control cells. The black dots represent the new commercial buildings, and the solid lines represent district borders of São Paulo. Panels B and C present the results of the matching procedure using the Proximity Probability Score for each treated group.

Table F1. Effects of New Commercial Buildings: Larger Sample of Neighborhoods

	0-250m				250-500m			
	Log Estabs (1)	Log Workers (2)	% College (3)	Wage Premium (4)	Log Estabs (5)	Log Workers (6)	% College (7)	Wage Premium (8)
Panel A. High-Skilled Offices								
α_r	0.1538***	0.3854***	0.0200	0.0936***	0.0494	0.1650**	-0.0060	-0.0005
	(0.0499)	(0.1025)	(0.0177)	(0.0318)	(0.0335)	(0.0697)	(0.0156)	(0.0171)
R^2	0.02382	0.03339	0.00215	0.02440	0.00233	0.00562	0.00003	-0.00012
Obs	3,608	3,543	3,543	3,543	8,167	7,978	7,978	7,978
Panel B. Low-Skilled Offices								
α_r	0.1029**	0.4228***	0.0345*	-0.0075	0.0698***	0.1411*	0.0060	0.0098
	(0.0465)	(0.1228)	(0.0206)	(0.0260)	(0.0263)	(0.0748)	(0.0115)	(0.0147)
R^2	0.01260	0.02924	0.00607	-0.00008	0.00654	0.00341	0.00008	0.00025
Obs	3,739	3,696	3,696	3,696	8,641	8,475	8,475	8,475
Panel C. Local Services								
α_r	0.0385	0.1653***	0.0314***	0.0266*	0.0094	0.0343	0.0032	0.0190*
	(0.0319)	(0.0595)	(0.0106)	(0.0155)	(0.0182)	(0.0374)	(0.0080)	(0.0106)
R^2	0.00306	0.01685	0.01630	0.00553	0.00009	0.00056	0.00003	0.00247
Obs	4,114	4,096	4,096	4,096	9,606	9,501	9,501	9,501
Panel D. Non-Offices								
α_r	0.0717*	0.2934***	0.0713***	0.0503***	0.0492**	0.1398***	0.0061	0.0104
	(0.0426)	(0.0747)	(0.0157)	(0.0190)	(0.0243)	(0.0510)	(0.0104)	(0.0133)
R^2	0.00631	0.02395	0.03936	0.01049	0.00334	0.00537	0.00015	0.00034
Obs	4,027	3,986	3,986	3,986	9,453	9,348	9,348	9,348

Notes: This table reports estimates of α_r in Equation (14) for different outcome variables using alternative samples of cells, as described in Section F. Standard errors clustered at the cell level are displayed in parentheses. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

Alternative Thresholds - New Buildings

In Table F2, 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, as indicated in the column headers.

Table F2. Effects of New Commercial Buildings: First-ring Cells

Variable	Baseline	n = 600	n = 400	sh = 35%	sh = 15%
High-skilled Offices					
Log Estabs	0.1598 (0.0547)	0.0966 (0.0655)	0.1266 (0.0514)	0.0176 (0.0613)	-0.004 (0.0613)
Log Workers	0.3123 (0.1134)	0.3302 (0.1322)	0.2287 (0.1021)	0.2465 (0.1215)	0.1612 (0.1142)
% College	0.0403 (0.0164)	0.0441 (0.021)	0.0495 (0.014)	0.0327 (0.0201)	0.0383 (0.0173)
Wage Premium	0.1167 (0.0394)	0.0602 (0.0516)	0.1055 (0.0296)	0.072 (0.0428)	0.0616 (0.0369)
Low-skilled Offices					
Log Estabs	0.0161 (0.0521)	0.046 (0.0717)	-0.0033 (0.0437)	0.0438 (0.0564)	0.0525 (0.0513)
Log Workers	0.145 (0.1616)	0.2003 (0.2176)	0.0717 (0.1633)	0.2079 (0.1789)	0.2076 (0.1605)
% College	-0.0272 (0.024)	-0.0135 (0.0297)	-0.0203 (0.0206)	-0.0194 (0.0254)	-0.0697 (0.0275)
Wage Premium	-0.0668 (0.0287)	-0.0692 (0.0378)	-0.0397 (0.0244)	-0.066 (0.03)	-0.0713 (0.0278)
Local Services					
Log Estabs	0.083 (0.0332)	0.1322 (0.0381)	0.0045 (0.0292)	0.061 (0.0355)	0.0825 (0.0338)
Log Workers	0.1846 (0.0575)	0.2189 (0.0647)	0.1543 (0.0495)	0.0893 (0.0707)	0.1597 (0.0569)
% College	0.0103 (0.0129)	0.0292 (0.0129)	0.0231 (0.0117)	-0.0056 (0.0121)	0.0041 (0.0128)
Wage Premium	3e-04 (0.0187)	0.0104 (0.0179)	0.0116 (0.0148)	-0.0072 (0.0161)	-0.0033 (0.0171)
Non-offices					
Log Estabs	-0.0106 (0.0476)	2e-04 (0.0568)	0.0277 (0.0412)	-9e-04 (0.0616)	-0.0431 (0.0461)
Log Workers	0.0906 (0.0777)	0.1329 (0.0849)	0.0589 (0.0696)	0.116 (0.1066)	0.0443 (0.0735)
% College	0.0354 (0.02)	0.0336 (0.0239)	0.1046 (0.0183)	0.0804 (0.0208)	0.0752 (0.0195)
Wage Premium	0.006 (0.0241)	0.0091 (0.0319)	0.0386 (0.022)	0.0636 (0.0292)	0.046 (0.0279)
Obs	1890	1290	2310	1650	1980

Notes: This table displays estimates of α_r in Equation (14) for various outcomes based on alternative thresholds to select new commercial buildings, as described in Section F. These results are exclusively for the effects of new commercial buildings on first ring cells. 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.

Estimation with No Matching

In Table F3, Columns (1) to (3) display pretreatment summary statistics of the outcome variables (mean and standard variation), and Columns (4) and (5) display the p-value of the difference in means between treated and control cells. Table F4 shows estimates of Equation (14) using the new sample, which replicates the structure of Table 3.

Table F3. Baseline Summary Statistics

Variable	Treated 1 (T1)	Treated 2 (T2)	Control (C)	t-test T1-C	t-test T2-C
Workers					
High-skilled Offices	243.0 (382.4)	159.4 (390.4)	154.7 (622.5)	[0.014]	[0.873]
Low-skilled Offices	162.7 (440.8)	225.8 (722.5)	137.4 (579.7)	[0.509]	[0.032]
Local Services	208.2 (215.3)	171.1 (233.4)	163.9 (402.8)	[0.039]	[0.699]
Non-offices	253.3 (337.5)	194.8 (297.1)	132.2 (240.1)	[0.000]	[0.000]
Establishments					
High-skilled Offices	14.8 (14.8)	10.7 (11.8)	7.4 (9.3)	[0.000]	[0.000]
Low-skilled Offices	14.4 (18.0)	12.6 (12.3)	9.8 (11.0)	[0.001]	[0.000]
Local Services	20.6 (14.4)	20.3 (15.4)	20.1 (34.2)	[0.733]	[0.885]
Non-offices	13.4 (10.7)	12.0 (9.2)	8.9 (7.1)	[0.000]	[0.000]
Wages					
High-skilled Offices	4937.6 (4076.7)	4061.8 (3227.0)	3358.7 (3171.4)	[0.000]	[0.000]
Low-skilled Offices	3325.2 (3062.4)	2492.8 (2043.8)	2060.3 (1272.2)	[0.000]	[0.000]
Local Services	2926.9 (2485.3)	2432.9 (1625.3)	2007.7 (1300.2)	[0.000]	[0.000]
Non-offices	4961.9 (3711.1)	4092.7 (2908.0)	3480.1 (2554.4)	[0.000]	[0.000]
% College					
High-skilled Offices	58.8 (24.0)	55.8 (21.6)	51.9 (23.3)	[0.000]	[0.004]
Low-skilled Offices	43.5 (22.1)	37.1 (20.9)	32.7 (19.7)	[0.000]	[0.000]
Local Services	29.7 (19.6)	25.0 (16.7)	21.7 (15.5)	[0.000]	[0.001]
Non-offices	39.7 (21.2)	41.0 (21.6)	37.4 (20.9)	[0.180]	[0.005]
Observations	63	138	255		

Notes: This table presents baseline summary statistics of treated and control groups using observations pre-treatment observations, i.e., prior to 2006. Standard deviations of variables appear in parentheses and p-values for differences of means appear in square brackets. Columns (1) and (2) show the mean and standard deviations for treated and control cells in the first sample, respectively, and Columns (4) and (5) do the same for the second sample. Columns (3 and (6) show the p-value of the t-tests of the difference in means in each case. Average wages are in 2017 reais.

Table F4. Effects of New Commercial Buildings: No Matching

	0-250m				250-500m			
	Log Estabs (1)	Log Workers (2)	% College (3)	Wage Premium (4)	Log Estabs (5)	Log Workers (6)	% College (7)	Wage Premium (8)
Panel A. High-Skilled Offices								
α_r	0.0701	0.3021***	0.0508***	0.0582	0.0771**	0.1771**	0.0137	-0.0013
	(0.0539)	(0.1170)	(0.0156)	(0.0368)	(0.0392)	(0.0879)	(0.0153)	(0.0199)
R^2	0.00401	0.01695	0.01294	0.00787	0.00650	0.00749	0.00103	-0.00016
Obs	4,770	4,770	4,770	4,770	5,895	5,895	5,895	5,895
Panel B. Low-Skilled Offices								
α_r	0.0339	0.0904	0.0009	-0.0685**	0.0831**	0.1150	0.0178	0.0179
	(0.0517)	(0.1630)	(0.0235)	(0.0288)	(0.0365)	(0.0966)	(0.0145)	(0.0198)
R^2	0.00104	0.00082	-0.00021	0.01469	0.00993	0.00238	0.00213	0.00129
Obs	4,770	4,770	4,770	4,770	5,895	5,895	5,895	5,895
Panel C. Local Services								
α_r	0.0410	0.1085*	-0.0054	-0.0267*	0.0028	0.0115	-0.0093	-0.0252**
	(0.0327)	(0.0566)	(0.0121)	(0.0160)	(0.0212)	(0.0411)	(0.0094)	(0.0119)
R^2	0.00325	0.00736	0.00017	0.00540	-0.00014	-0.00005	0.00135	0.00655
Obs	4,770	4,770	4,770	4,770	5,895	5,895	5,895	5,895
Panel D. Non-Offices								
α_r	-0.0566	-0.0250	0.0607***	0.0148	-0.0355	-0.0919	-0.0046	-0.0238
	(0.0476)	(0.0728)	(0.0196)	(0.0239)	(0.0300)	(0.0597)	(0.0131)	(0.0176)
R^2	0.00297	-0.00005	0.02015	0.00048	0.00175	0.00286	0.00000001	0.00227
Obs	4,770	4,770	4,770	4,770	5,895	5,895	5,895	5,895

Notes: This table reports estimates of new building effects for different outcome variables without employing matching, as described in Section F. Standard errors clustered at the cell level are displayed in parentheses. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

Effects of New Residential Buildings

To estimate the effects of large residential buildings on economic activity, I use the IPTU dataset to select the top 100 residential buildings in terms of built area inaugurated between 2006 and 2013. Because I do not have data on residents for each building, the built area is the most reliable piece of information available to measure the magnitude of the shock. Table F5 shows that these buildings are, on average, larger than the sample of commercial buildings, although there is less variation in this sample (see Table 1 for a comparison).[46]

Then, I repeat the procedure described in Section 4.2 to select the neighborhoods for the analysis and to construct the treatment variables. Figure F2 shows the location of these buildings, together with the treatment and control cells. Compared to the sample of commercial buildings, they are more spread throughout the city and farther from high employment areas..

Table F6 presents the results from Equation (14) in this new setting for the same sectoral variables examined so far. A few coefficients related to local services and high-skilled offices are negative and significant, which might indicate a mild level of spurious correlation. Nonetheless, on the whole, the results indicate no meaningful effects on economic activity in nearby neighborhoods, particularly on local services.

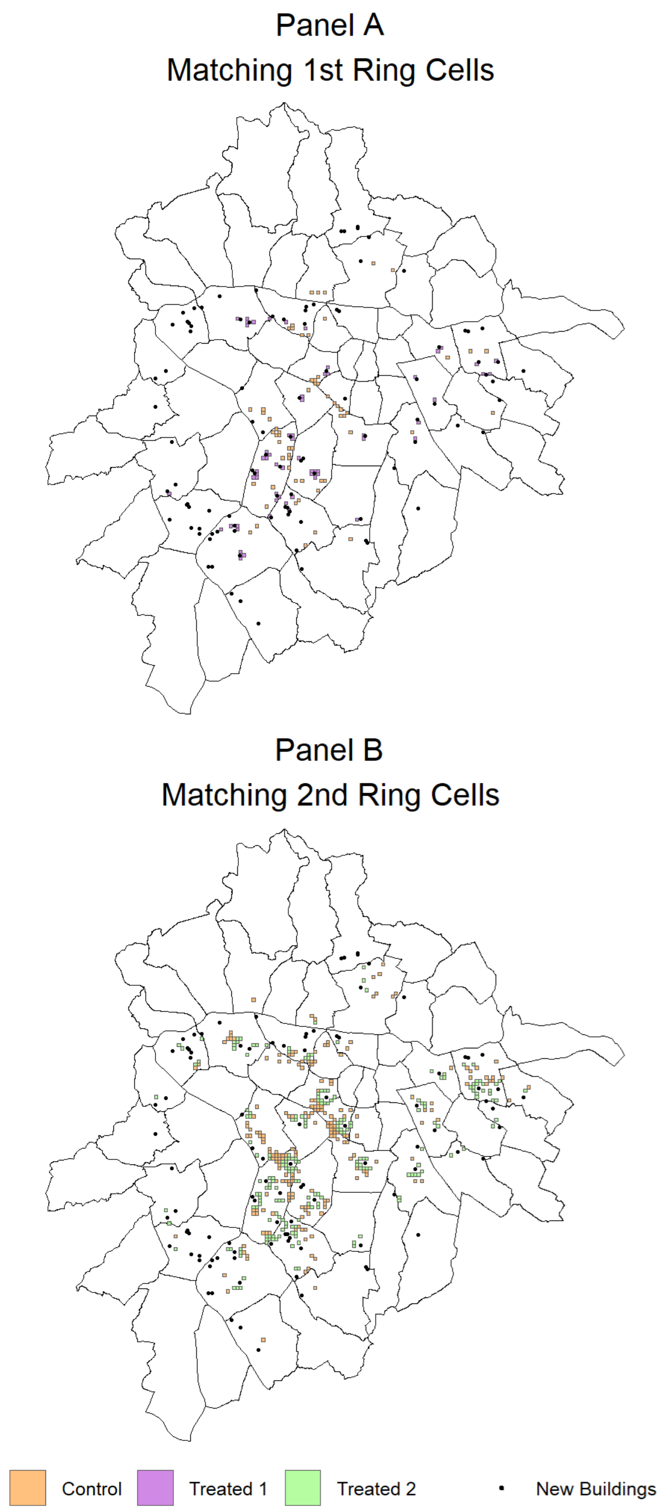
Table F5. Summary Statistics: New Residential Buildings

	Median	Mean	Std. Deviation	Min	Max
Total Land Area (m^2)	$8,033.5$	$9,245.1$	$4,734.7$	$3,485$	$26,620$
Occupied Land Area (m^2)	$5,030$	$5,760.7$	$2,885.1$	$1,434$	$15,959$
Built-Area-Ratio	9.4	10.9	4.9	4.6	30.5

Notes: This table displays summary statistics for the sample of new residential buildings.

[46]In my sample of commercial buildings, the correlation between built area and average employment is 0.93.

Figure F2. Empirical Analysis Setup



Notes: This figure depicts the design of the empirical analysis using residential buildings as the shock. Details are provided in Section F.

Table F6. Effects of New Residential Buildings

	0-250m				250-500m			
	Log Estabs (1)	Log Workers (2)	% College (3)	Wage Premium (4)	Log Estabs (5)	Log Workers (6)	% College (7)	Wage Premium (8)
Panel A. High-Skilled Offices								
α_r	-0.0506	-0.1615	0.0026	0.0399	-0.0884***	-0.1186	0.0103	-0.0325*
	(0.0667)	(0.1614)	(0.0295)	(0.0496)	(0.0337)	(0.0726)	(0.0115)	(0.0190)
R^2	0.00289	0.00636	-0.00050	0.00434	0.00947	0.00387	0.00069	0.00415
Obs	1,830	1,830	1,830	1,830	6,180	6,180	6,180	6,180
Panel B. Low-Skilled Offices								
α_r	0.0330	-0.1384	-0.0499**	-0.0712**	-0.0020	0.1949**	0.0111	0.0377**
	(0.0569)	(0.1374)	(0.0254)	(0.0358)	(0.0339)	(0.0934)	(0.0124)	(0.0157)
R^2	0.00150	0.00419	0.01859	0.02190	-0.00016	0.00657	0.00082	0.00725
Obs	1,830	1,830	1,830	1,830	6,180	6,180	6,180	6,180
Panel C. Local Services								
α_r	0.0658	0.0736	-0.0317**	0.0066	-0.0278	-0.0294	-0.0068	-0.0077
	(0.0577)	(0.0917)	(0.0158)	(0.0235)	(0.0207)	(0.0342)	(0.0082)	(0.0122)
R^2	0.01015	0.00388	0.01897	-0.00011	0.00230	0.00076	0.00080	0.00049
Obs	1,830	1,830	1,830	1,830	6,180	6,180	6,180	6,180
Panel D. Non-Offices								
α_r	0.0577	0.0462	0.0554***	0.0386	-0.0517**	-0.2065***	-0.0039	-0.0235
	(0.0607)	(0.1120)	(0.0205)	(0.0314)	(0.0261)	(0.0554)	(0.0103)	(0.0168)
R^2	0.00539	0.00040	0.03130	0.00662	0.00465	0.01730	0.000001	0.00269
Obs	1,830	1,830	1,830	1,830	6,180	6,180	6,180	6,180

Notes: This table reports estimates of the effects of new residential building 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.

Continuous Treatment

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

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

where $T_{c,t}$ is the new treatment variable based on the closest new building used to separate neighborhoods into treatment and control groups (see Section 4.2). I explore a linear and an exponential treatment variable:

$$T^{lin}_{c,t} = (1 - dist) \times \mathbb{1}(Treated_t) \tag{F.2}$$

$$T^{exp}_{c,t} = \exp(1 - dist) \times \mathbb{1}(Treated_t) \quad , \tag{F.3}$$

where $dist \in [0, 1]$ is the distance, in kilometers, between the cell's centroid and the closest new building, and $\mathbb{1}(Treated_t)$ is an indicator of treatment in period t, 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. Note that given this definition, α is expected to be positive.

Table F7. Effects of New Commercial Buildings: Linear Continuous Treatment

	Log Estabs (1)	Log Workers (2)	% College (3)	Wage Premium (4)
Panel A. High-Skilled Offices				
$(1 - dist) \times \mathbb{1}(Treated_t)$	0.1597***	0.2902***	0.0505***	0.1409***
	(0.0444)	(0.0964)	(0.0157)	(0.0259)
R^2	0.01108	0.00969	0.00592	0.03012
Observations	5,235	5,235	5,235	5,235
Panel B. Low-Skilled Offices				
$(1 - dist) \times \mathbb{1}(Treated_t)$	0.0974**	-0.0958	-0.0397**	0.0771***
	(0.0404)	(0.1175)	(0.0176)	(0.0234)
R^2	0.00317	0.00043	0.00374	-0.00143
Observations	5,235	5,235	5,235	5,235
Panel C. Local Services				
$(1 - dist) \times \mathbb{1}(Treated_t)$	0.0925***	0.2556***	-0.0186*	0.0377***
	(0.0258)	(0.0438)	(0.0104)	(0.0144)
R^2	0.00773	0.01419	0.00365	-0.00250
Observations	5,235	5,235	5,235	5,235
Panel D. Non-Offices				
$(1 - dist) \times \mathbb{1}(Treated_t)$	-0.0814**	-0.2356***	0.0603***	0.0776***
	(0.0360)	(0.0665)	(0.0150)	(0.0196)
R^2	0.00251	0.00728	0.00801	-0.00181
Observations	5,235	5,235	5,235	5,235

Notes: This table reports estimates of new building effects for different outcome variables considering a linear continuous treatment variable according to Equations (F.1) and (F.2). Standard errors clustered at the cell level are displayed in parentheses. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

Table F8. Effects of New Commercial Buildings: Exponential Continuous Treatment

	Log Estabs (1)	Log Workers (2)	% College (3)	Wage Premium (4)
Panel A. High-Skilled Offices				
$\exp(1 - dist) \times \mathbb{1}(Treated_t)$	0.1229***	0.2058***	0.0406***	0.1048***
	(0.0344)	(0.0767)	(0.0125)	(0.0201)
R^2	0.01042	0.00694	0.00685	0.02508
Observations	5,235	5,235	5,235	5,235
Panel B. Low-Skilled Offices				
$\exp(1 - dist) \times \mathbb{1}(Treated_t)$	0.0812***	-0.0721	-0.0316**	0.0708***
	(0.0308)	(0.0896)	(0.0135)	(0.0178)
R^2	0.00485	0.00034	0.00413	0.00615
Observations	5,235	5,235	5,235	5,235
Panel C. Local Services				
$\exp(1 - dist) \times \mathbb{1}(Treated_t)$	0.0748***	0.2141***	-0.0119	0.0365***
	(0.0210)	(0.0356)	(0.0081)	(0.0113)
R^2	0.00925	0.02147	0.00197	0.00329
Observations	5,235	5,235	5,235	5,235
Panel D. Non-Offices				
$\exp(1 - dist) \times \mathbb{1}(Treated_t)$	-0.0663**	-0.1798***	0.0498***	0.0729***
	(0.0283)	(0.0548)	(0.0118)	(0.0156)
R^2	0.00324	0.00655	0.01100	0.00779
Observations	5,235	5,235	5,235	5,235

Notes: This table reports estimates of new building effects for different outcome variables considering an exponential continuous treatment variable according to Equations (F.1) and (F.3). Standard errors clustered at the cell level are displayed in parentheses. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

SE Clustering using Closest New Building

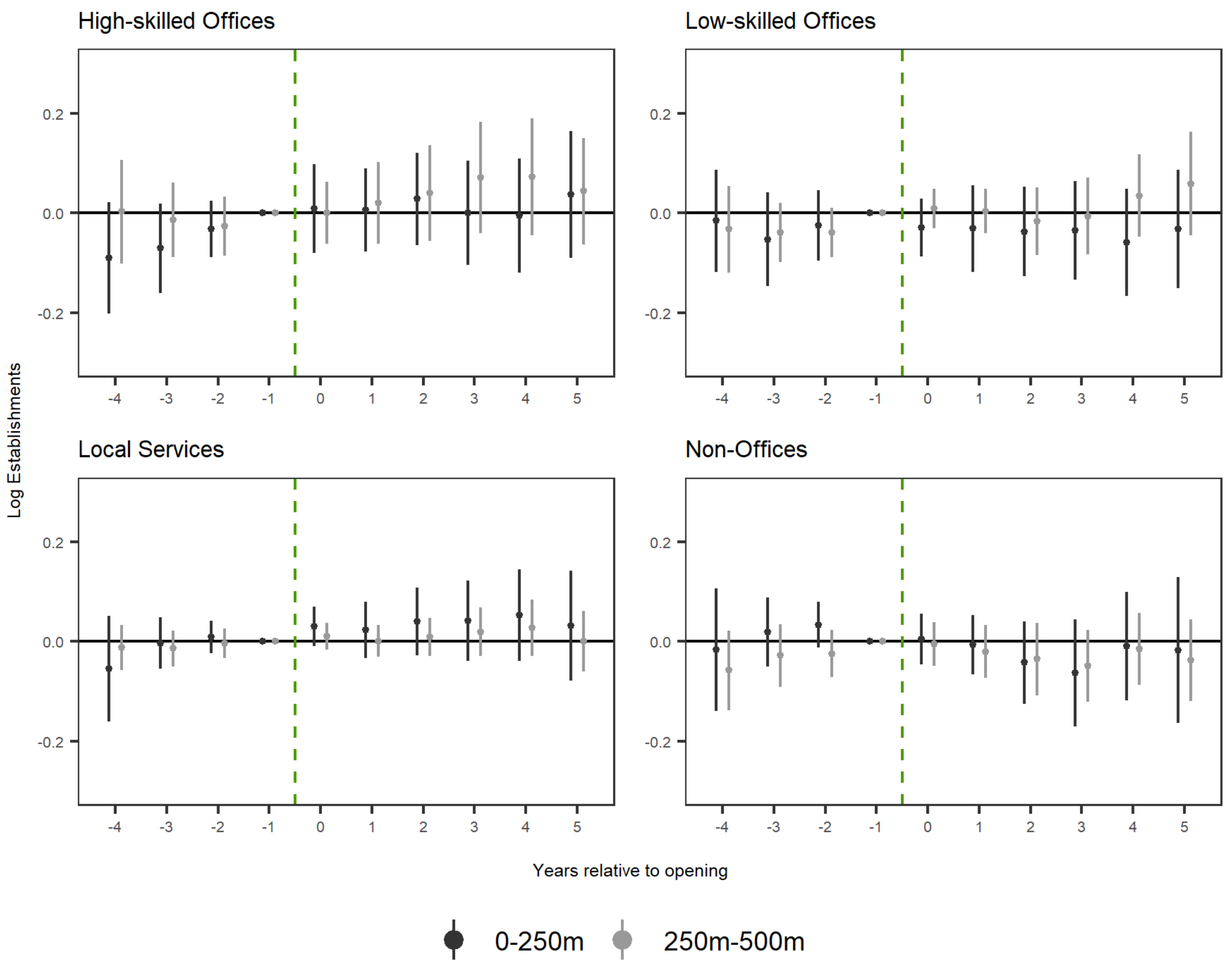
Table F9. Effects of New Commercial Buildings: Alternative Clustering

	0-250m				250-500m			
	Log Estabs (1)	Log Workers (2)	% College (3)	Wage Premium (4)	Log Estabs (5)	Log Workers (6)	% College (7)	Wage Premium (8)
Panel A. High-Skilled Offices								
α_r	0.1598**	0.3123***	0.0403**	0.1167***	0.0409	0.1611	0.0007	0.0286
	(0.0621)	(0.1052)	(0.0196)	(0.0448)	(0.0499)	(0.1433)	(0.0181)	(0.0256)
R^2	0.03736	0.03239	0.01903	0.04969	0.00199	0.00667	-0.00024	0.00339
Obs	1,890	1,890	1,890	1,890	4,140	4,140	4,140	4,140
Panel B. Low-Skilled Offices								
α_r	0.0161	0.1450	-0.0272	-0.0668**	0.0674	0.0862	-0.0137	0.0342
	(0.0614)	(0.1531)	(0.0282)	(0.0319)	(0.0476)	(0.1138)	(0.0192)	(0.0232)
R^2	-0.00004	0.00347	0.00537	0.02150	0.00780	0.00142	0.00139	0.00557
Obs	1,890	1,890	1,890	1,890	4,140	4,140	4,140	4,140
Panel C. Local Services								
α_r	0.0830***	0.1846***	0.0103	0.0003	0.0261	0.0312	-0.0044	0.0027
	(0.0296)	(0.0524)	(0.0146)	(0.0243)	(0.0327)	(0.0649)	(0.0104)	(0.0183)
R^2	0.02035	0.03724	0.00169	-0.00053	0.00235	0.00079	0.00017	-0.00016
Obs	1,890	1,890	1,890	1,890	4,140	4,140	4,140	4,140
Panel D. Non-Offices								
α_r	-0.0106	0.0906	0.0354	0.0060	-0.0005	-0.0596	-0.0045	-0.0300*
	(0.0502)	(0.0855)	(0.0236)	(0.0309)	(0.0286)	(0.0873)	(0.0122)	(0.0162)
R^2	-0.00032	0.00370	0.01203	-0.00031	-0.00024	0.00127	-0.00004	0.00426
Obs	1,890	1,890	1,890	1,890	4,140	4,140	4,140	4,140

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

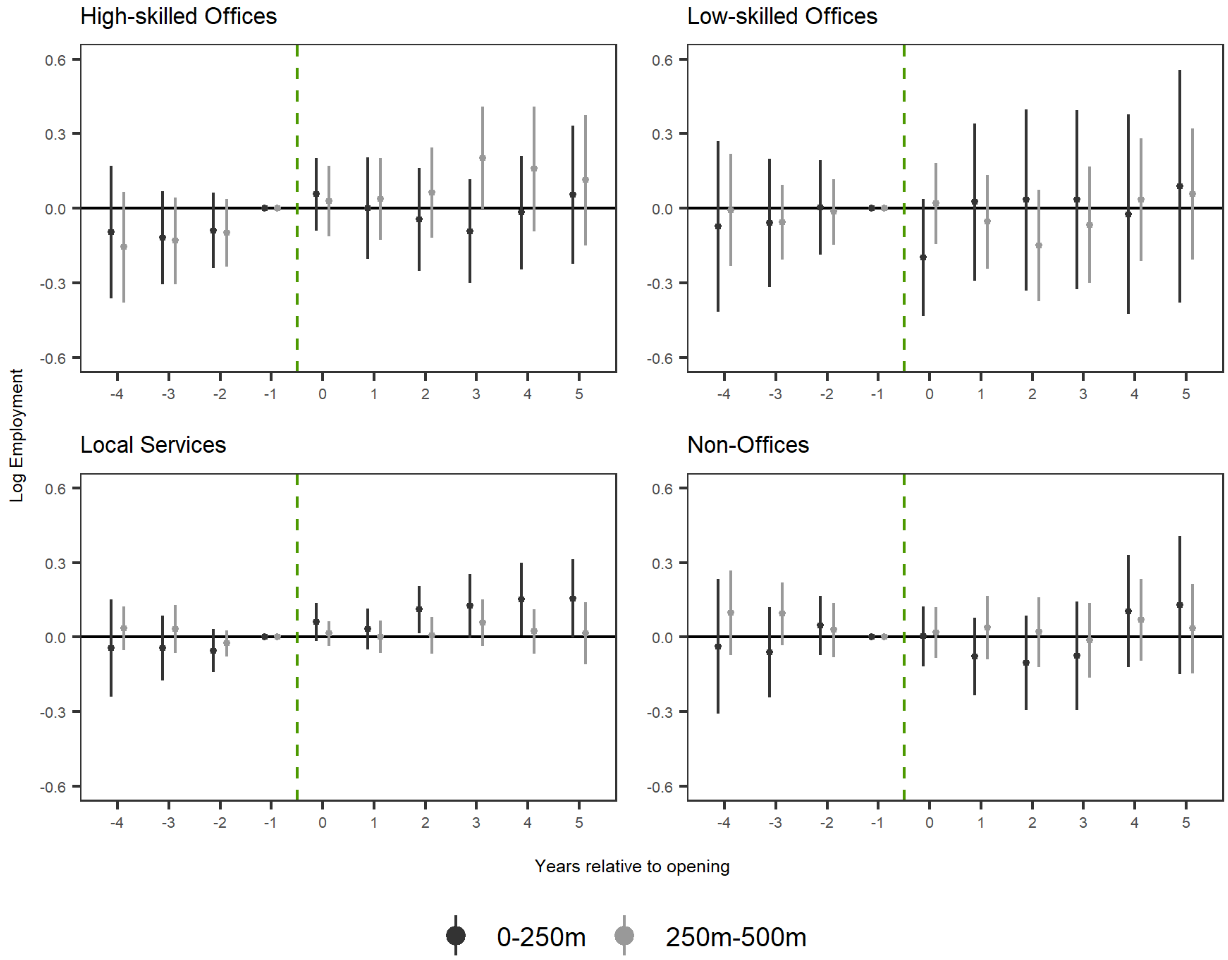
Callaway and Sant'Anna (2021)'s Estimator

Figure F3. Event Study using Callaway and Sant'Anna (2021): Number of Establishments by Sector



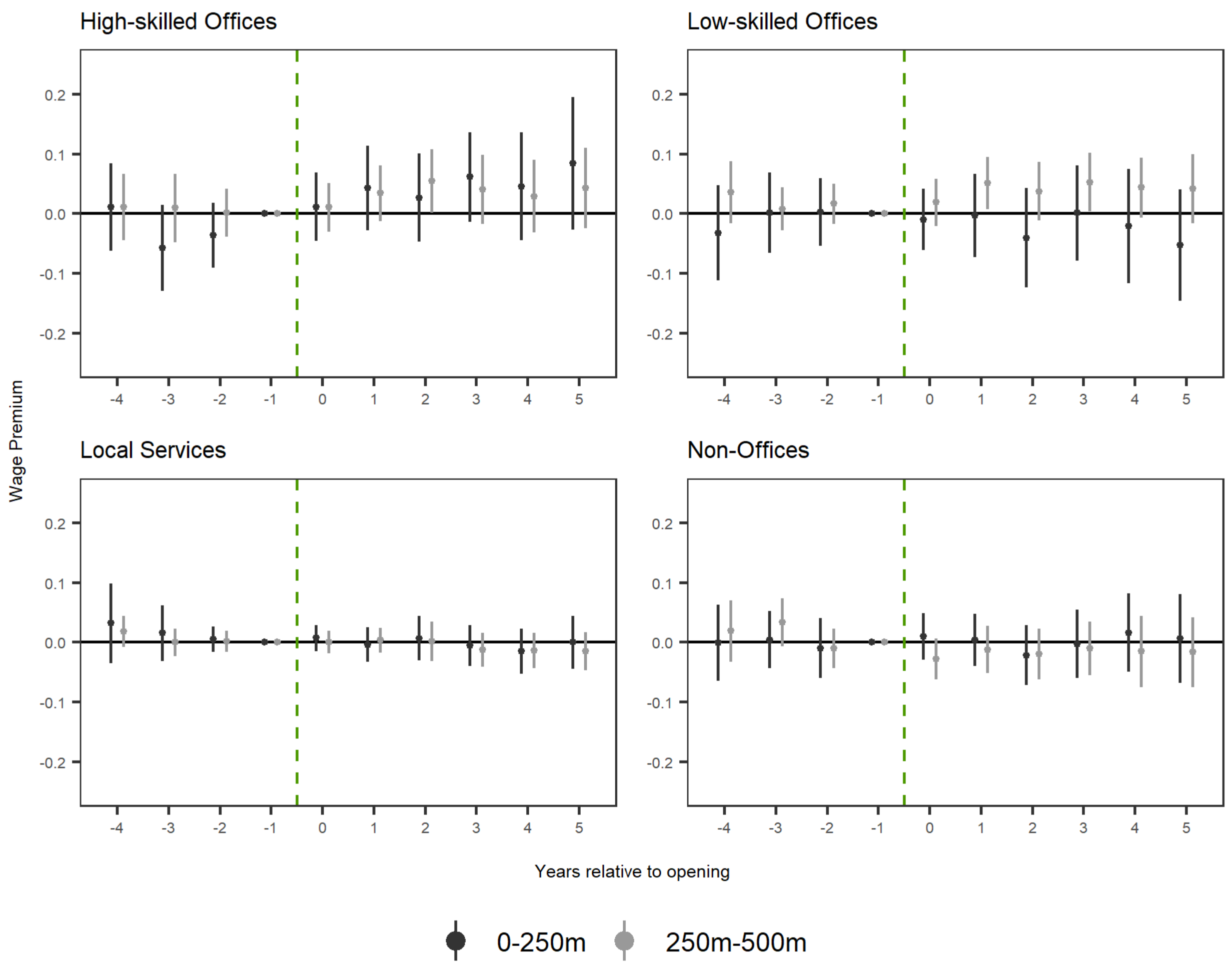
Note: This figure plots coefficients from Callaway and Sant'Anna (2021)'s estimator using the log number of establishments for different sectors as the outcome variable. The definition of each sector is described in Section 2.2. The bars indicates the 95% confidence interval, where standard errors are clustered at the cell level.

Figure F4. Event Study using Callaway and Sant'Anna (2021): Employment by Sector



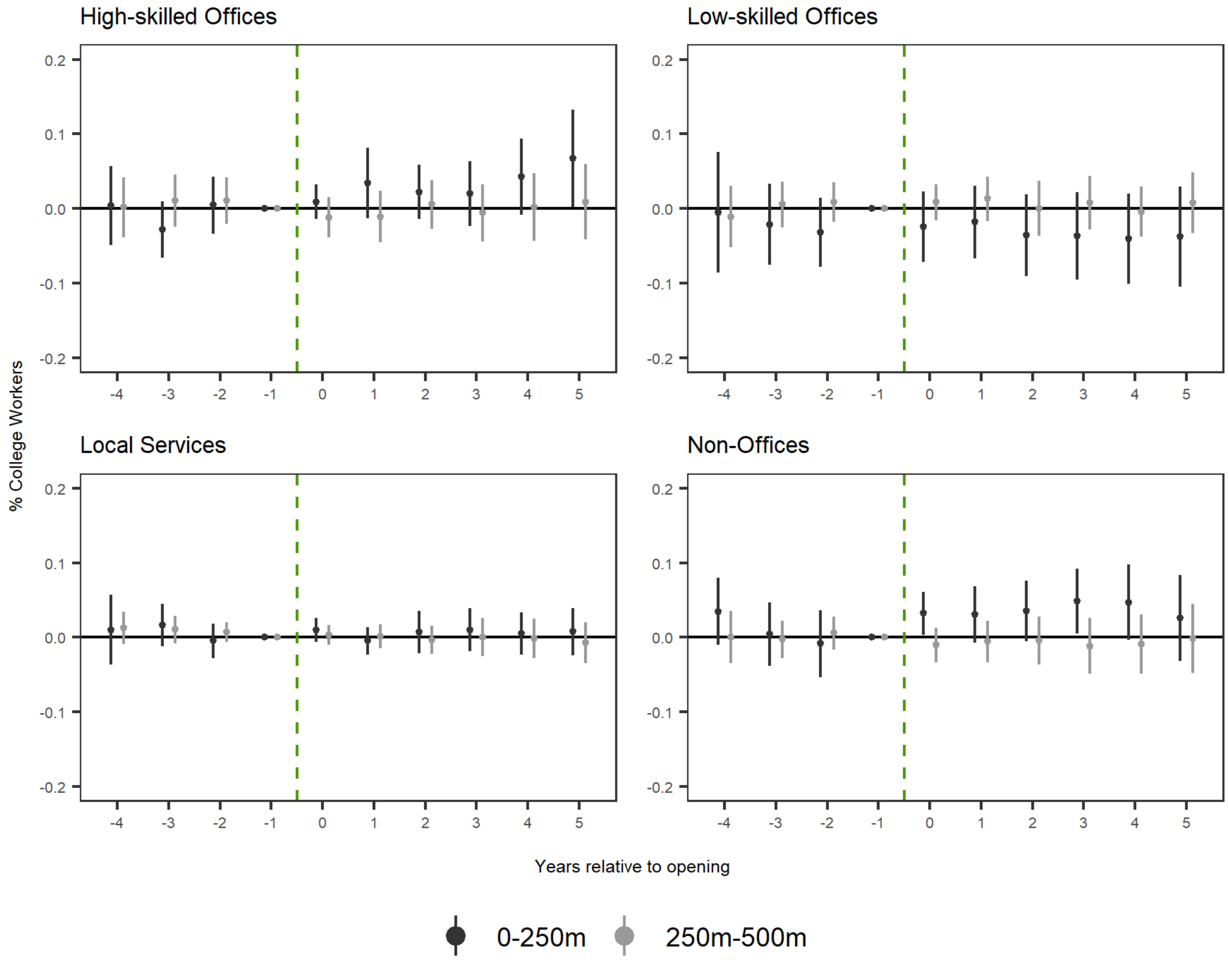
Note: This figure plots coefficients from Callaway and Sant'Anna (2021)'s estimator using log employment for different sectors as the outcome variable. The definition of each sector is described in Section 2.2. The bars indicates the 95% confidence interval, where standard errors are clustered at the cell level.

Figure F5. Event Study using Callaway and Sant'Anna (2021): Wage Premium by Sector



Note: This figure plots coefficients from Callaway and Sant'Anna (2021)'s estimator using mean establishment wage premium (weighted by establishment size) for different sectors as the outcome variable. 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 F6. Event Study using Callaway and Sant'Anna (2021): Share of College Workers by Sector



Note: This figure plots coefficients from Callaway and Sant'Anna (2021)'s estimator using the share of workers with college degree for different sectors as the outcome variable. 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.