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

Thiago Patto†

November 6, 2023

Click here for latest version

**Abstract**

I explore the opening of large commercial buildings to study how local economic activity is affected by an employment shock. My methodology combines the typical ring approach, which involves comparing neighborhoods nearer and farther away, with a matching procedure to obtain samples of neighborhoods with a similar probability of observing a new building in their vicinity. I find that neighborhoods within 250 meters of a new building experience a 12.9% differential increase in employment, driven by high-skilled offices and local services. I estimate that for every two additional jobs created in high-skilled offices, one job is created in local services. I also present suggestive evidence that the productivity of high-skilled offices is affected. There is an increase in the share of college-educated workers and the wages in this sector, which seems to be driven to a good extent by the entry of new firms. Overall, my findings indicate that both productivity spillovers and local demand are crucial ingredients of urban concentration. New buildings increase the productivity of high-skilled offices nearby, attracting more firms in this sector and raising the demand for non-tradable goods provided by local services.

*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.

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

# 1 Introduction

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

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

Using the opening of new large commercial buildings in São Paulo as a natural experiment, this article helps to fill this gap by studying how a sudden increase in urban concentration in one location impacts neighborhoods in the vicinity. The analysis focuses on how different sectors respond to these shocks and compares neighborhoods more or less exposed to these buildings to identify the effects. My sector classification aims to differentiate be-

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

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

tween establishments that produce non-tradable goods and, therefore, are more dependent on local demand (local services) and those that produce tradable goods (high- and low-skilled offices).[3]

With 21.7 million inhabitants, São Paulo is one of the ten largest metropolitan areas globally and the largest in the Americas. It has almost doubled its population in 50 years, thus providing a compelling setting for the study.[4] Furthermore, it contains high-quality data with detailed location information. I use two administrative sources that can be accurately geocoded to perform the empirical analysis: property tax records from the 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 buildings recently inaugurated 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 develop a difference-in-differences model with staggered treatment adoption. I define treated and control neighborhoods based on the nearest building site and the year of treatment based on the first building within the range that characterizes that neighborhood as a treated unit. These definitions are necessary since some neighborhoods are close to more than one building.[5]

A central challenge in this empirical approach is that developers endogenously choose buildings' locations. I restrict the sample to cells within 1 km of a new building and explore small variations in distance to attenuate concerns about selection, as treated and control neighborhoods belong to the same regions and are presumably similar in several dimensions. I

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

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

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

also take advantage of the panel structure of the data and estimate event-study specifications to check for pre-trends.

Nevertheless, it is still possible that neighborhoods face local shocks correlated with the likelihood of being close to a new building. To address this concern, I build on Qian and Tan (2021) and develop a procedure to compare neighborhoods with a similar probability of being treated. I estimate a propensity score model that uses neighborhood characteristics prior to buildings' inaugurations to predict which locations are more likely to receive a new construction. From the predicted values, I create a measure that expresses the ex-ante probability that a cell observes a new building in its vicinity. Then, based on this measure, I 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 an 12.9% 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 14.3% and 19.9%, respectively. Considering the average employment one period prior to the treatment, it means that approximately one job is created in local services for each additional two jobs created in high-skilled offices.

I also show evidence consistent with the idea that the productivity of high-skilled offices is affected in treated neighborhoods. There is a differential increase of 3 percentage points in the share of college-educated workers and a differential increase of 5.4% in the average wage premium paid by this sector.[6] However, additional evidence suggests that these effects are driven to a good extent by changes in firm composition, meaning that the sorting of high-wage firms in treated neighborhoods is relevant to understanding the findings.

To interpret the results, I propose a stylized model of firm location choice built on Ahlfeldt et al. (2015) that yields an equilibrium distribution across neighborhoods. In the model, firms

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

belong to different sectors and produce goods whose prices are defined by the larger economy, with the exception of local services, whose price is defined at the neighborhood level. Each pair neighborhood-sector has a local TFP, treated in principle as fixed. Because I assume that workers in a given neighborhood spend a fraction of their income on local services, this sector responds to changes in local economic activity driven by other sectors. I demonstrate that when there is a shift in the TFP of one sector in one neighborhood, this sector expands its presence in the neighborhood, leading to an increase in wages and employment.[7] Local services are then indirectly benefited and also expand. Thus, the estimated growth ratio of 0.5 between local services and offices can be interpreted as a multiplier effect.

I also explore the impacts of new buildings on other dimensions. I show that local services and offices expand at the extensive margin in treated neighborhoods, with a differential increase of around 8% in the number of establishments for each sector. Regarding the supply of floor space, I find no evidence of an impact on treated neighborhoods, suggesting that the increase in the number of establishments occurs through a decrease in vacancy rates.

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

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

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

with the existence of productivity spillovers at the neighborhood level.

In this sense, my findings are closely related and complementary to Baum-Snow et al. (2021). They estimate productivity externalities at a similar scale using a peer effects model and firm revenue as a measure of productivity. While their approach delivers the estimation of structural parameters used to run counterfactuals, it restricts the analysis to high-skilled services. This paper, on the other hand, provides a broader description of the process of urban concentration.

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

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

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

# 2 Data and Descriptive Evidence

## 2.1 Data Sources

RAIS covers Brazil's formal labor market, with 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 latter two to compute monthly wages. An essential piece of information for this study is the establishments' complete addresses, which employers report annually.

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

I use RAIS to create a panel of private sector 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 and use these data to estimate separately, for each year, wage premiums for each establishment according to the following expression[9]

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

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

[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.

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

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

## 2.2 Sector Classification

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

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

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

[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.

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

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

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

## 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: a) some sectors are more concentrated than others, and b) more productive firms tend to be more concentrated. I focus on the city's districts where the empirical analysis is performed.

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

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 Luis Carlos Berrini (at 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] On the other hand, local services spread relatively more across 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 sectors. Businesses like restaurants and retail require proximity to potential customers, so they prefer to locate where people live or circulate. Thus, even though high-density

[15]This index is a more general version of the standard spatial Herfindahl index. It is given by $H = \boldsymbol{s'Ws}$, where $\boldsymbol{s}$ is a vector containing the share of firms in each neighborhood for a given sector, and $\boldsymbol{W}$ 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 neighborhoods $i$ and $j$, and $\bar{r}$ is the average distance across all neighborhoods. Guimaraes et al. (2011) argue that their measure has the advantage of considering both the concentration level within 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 Avenida Paulista. The traditional mansions were demolished to clear space for commercial buildings, which hosted the headquarters of many large companies. Avenida Brigadeiro Faria Lima began to gain relevance in the 1970s, and by the 1990s, it was already one of the densest employment areas. Nowadays, it is considered the most important business center of São Paulo, especially for industries such as finance and tech. Employment in Avenida Luis Carlos Berrini followed the same trend as Faria Lima but with a few years of delay

neighborhoods tend to have a larger provision of local services, firms in this sector also have some incentives to locate in low-density neighborhoods. On the other hand, firms in the financial or tech sector do not rely on local demand but may benefit from productivity spillovers in high-density locations. Depending on the magnitude of these spillovers, it can discourage firms from locating in isolated neighborhoods.

Regarding the second point, Figure 2 displays, for 2010, the distribution of establishments in high-skilled offices for different terciles of establishment wage premium. It is worth noting that establishments in the first tercile are more dispersed than in the third tercile, i.e., the concentration level increases with the wage premium. To show that this fact is not entirely driven by industry composition, in Figure C1 I present the same analysis focusing on one specific industry, namely finance, with qualitatively similar results. These figures suggest that more productive firms have a stronger preference for concentration, possibly because their productivity is more sensitive to agglomeration effects.

These pieces of evidence underline that the preference to locate in high-density neighborhoods varies within and between sectors. In particular, the patterns observed for local services and offices seem to reflect different incentives for settling 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 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. I focus my analysis on changes in local productivity, as I argue that this particular channel is key to understanding the process of urban concentration. I consider a variation in local TFP for a specific sector to obtain testable predictions and compare them with the empirical evidence.

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

## 3.1 A Model of Firm Location Choice

Consider a city with multiple discrete neighborhoods indexed by $n$ and a continuum of firms in discrete sectors $\overline{E}_s$ choosing where to locate.[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 $\ell$ and floor space $f$ to produce a a homogeneous good within sectors. While rent prices $r_n$ are taken as given, each firm faces an upward-sloping labor supply curve and hence needs to choose the optimal wage $w_{e,s,n}$. To make it simple, I assume that the labor supply curve is identical for all firms in the same pair $(s, n)$, and since they solve the same profit maximization problem,

[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.

in practice, wages will be sector-neighborhood specific, i.e., $w_{e,s,n} = w_{s,n}\ \forall e$.

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

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

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

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

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

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

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

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

$n$ is given by

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

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

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

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

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

Expression (5) tells us that the relative sectoral presence in a neighborhood depends positively on local TFP and the firm commuter market access, and negatively on rent and local services prices. Naturally, higher benefits will be counterbalanced in equilibrium by higher costs. However, it is worth noting that while prices faced by firms in a given neighborhood are the same, TFP can vary by sector and thus generate differences in their spatial distribution.[18]

For local services, the same procedure yields

[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 $\beta$. However, while this channel could explain why some sectors are more likely to locate in low-rent price areas, it cannot account for differences in the concentration level of different sectors.

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

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

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

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

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

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

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

The second market clearing condition equates supply and demand for floor space in each neighborhood. Here, I assume that there is a competitive construction sector that uses land and capital as inputs. Land available in neighborhood $n$ is fixed at $\overline{T}_n$. Assuming a Cobb-Douglas production function in which the share of expenditures on land is $\alpha$, the floor space supply curve in neighborhood $n$ can be written as $\frac{\overline{T}_n}{\alpha} r_n^{\frac{1-\alpha}{\alpha}}$. Equating this expression with floor space demand yields[19]

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

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

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

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

where $K = \frac{\alpha}{\delta}\left(\frac{\chi}{\varepsilon}\right)^{1-\varepsilon} \beta^{\chi-\varepsilon}$. Since $\chi - \delta\varepsilon > 0$, it is clear from Equations (8) and (10) that higher productivity is associated with higher prices of local services and floor space, as expected.

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 $\Phi_s$ 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 B ■

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

Thus, there is an increase in the concentration of firms of sector $s$ in neighborhood $n$

that occurs due to a combination of heterogeneity in local productivity across sectors and common inputs, namely floor space and labor, whose price is locally defined. These inputs work as congestion forces that 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 $\dfrac{\partial \Phi_{s'}}{\partial A_{s,n}} \approx 0 \; \forall \; s'$, nominal wages in sector $s$ and local services increase, whereas it decreases for all other office sectors. Real wages (and employment) in sector $s$ and local services tend to increase as well if $\eta$ is not too high and, particularly for sector $s$ if the elasticity of $p_{LS,n}$ with respect to $A_{s,n}$ is not close to 1.*

**Proof.** See Appendix B ■

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

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

## 3.2 Spatial Sorting

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

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

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

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

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

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

# 4 Empirical Strategy

The empirical analysis based on the opening of large commercial buildings is grounded on the notion that a surge in employment in one neighborhood potentially causes a sizable shift in the TFP of nearby neighborhoods through spillover effects. In light of the model, this entails assuming that $A_{sn}$ is a function of the employment in adjacent locations, thereby making it partially endogenous. When a large building opens in a neighborhood $n$, some sectors in neighborhood $n'$ may experience a sudden increase $\Delta A_{sn'}$ in local TFP, affecting its local equilibrium in terms of firm composition. If this is true, the predictions obtained in the previous section should be consistent with the empirical findings. Additionally, the model can also assist the discussion about other channels being at play, such as demand effects and changes in amenities.

In order to obtain estimates with a causal interpretation, a key identification challenge is to distinguish the specific impact of a new building from the more general causes that may have attracted the building to a particular site in the first place, given that developers endogenously choose where to construct. From a broader perspective, new constructions are essentially part of the gradual development of a city.

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

My 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. By doing so, I estimate treatment (distance) effects from a sample of neighborhoods with a similar likelihood of observing a new building nearby.[20] The panel structure is also a key

[20] 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.

element in my analysis. Observing neighborhoods before and after an inauguration makes it possible to difference out invariant factors that influence local economic activity. It also allows me to employ event-study regressions to verify the existence of pre-trends.

## 4.1 Selecting "Treatment" Buildings

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

The next step is to merge RAIS and IPTU data using the address information to identify shocks in local employment. 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.[21]

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

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

[21] The IPTU data is not suitable on its own to identify shocks in local employment for two reasons. First, they do not provide good information about the buildings' inauguration year. In principle, one could use either the construction year information or the first appearance in the panel for this purpose, but there are significant inconsistencies in both sources of information. In particular, new constructions usually appear in the IPTU data one or two years after their actual conclusion. Secondly, some buildings may have a negligible employment impact if firms do not occupy them.

I define the inauguration year as the first period in which a building has 50 or more formal labor contracts in effect. This threshold is achieved in most cases in the first year after employment begins to kick in. Figure C2 shows that the occupation rate of these buildings accelerates rapidly after the inauguration.[22]

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

After applying these filters, I am left with 43 new commercial buildings, summarized in Table 1. Note that the average employment in these buildings ranges from a few more than 500 workers to almost 12,000. Regarding the number of establishments, there are buildings in which one big company settles, but in general, multiple establishments occupy the newly available space.[23]

Figure 3 shows the buildings' location and the variation in employment between 2005 and 2015. It is worth noting that employment has increased relatively more during this period in regions in the vicinity of the buildings, which reinforces the possibility that their location correlates with trends in local employment. Together with Figure 1, it also shows that these buildings tend to be located close to high-density neighborhoods, suggesting that the relative spatial distribution of employment has not changed dramatically. Another aspect to note is that many neighborhoods observe multiple new buildings nearby. This fact creates a few challenges when building the econometric specification, which are addressed in the next section.

In order to validate the sample obtained, I use Google Maps to find the buildings and take screenshots of them. Some of these pictures are shown in Figure C3. Since the website

[22] 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 sector.

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

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

## 4.2 Selecting Treated and Control Neighborhoods, Defining Treatment

After establishing the buildings to be used as local shocks, I define the sample of neighborhoods for the estimation. Departing from the grid that covers São Paulo, I first select cells whose centroid is within 1 km of at least one new building's site. Then, I exclude cells that received any of the 43 new commercial buildings, as the goal is to explore a shock external to the local neighborhood equilibrium. To focus on changes at the intensive margin and guarantee that treatment effects across different sectors are comparable, I also restrict the analysis to cells containing at least one worker in each sector in all periods. Thus, I obtain an initial sample of 478 cells between 2003 and 2017.[24]

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

- Treated Group 1 - first-ring cells: closest new building is within 0 to 250 meters;
- Treated Group 2 - second-ring cells: closest new building is within 250 to 500 meters ;

[24] As a robustness check, I provide alternative results relaxing the last restriction, reported in Table D1. Section 6.1 gives more details.

- 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. For example, suppose a cell is close to three new commercial buildings: one inaugurated in 2008 that is 350 meters distant, one inaugurated in 2009 that is 140 meters distant and one inaugurated in 2012 that is 60 meters distant. In this case, the closest new building is within 0 to 250 meters; therefore, the cell is in Treated Group 1. The first treatment period is 2009 because it is the inauguration year of the first building within 0 and 250 meters. Figure 4 panel A shows the spatial distribution of the 478 cells with the classification proposed. The black dots represent the new commercial buildings, and the solid lines represent district borders of São Paulo.

These definitions imply that my empirical design is not meant to identify the effects of one extra commercial building, but instead 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 (See Table 1) and may not be fully occupied in the very short term (See Figure C2). Thus, the interpretation would not be straightforward even in an ideal scenario in which buildings are adequately distant from each other. Additionally, the relative changes across sectors also offer information whose value depends less on the characterization of the shock. In particular, the relative growth between local services and offices has a precise definition, as discussed in Section 3.[25]

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

[25] It is also important to stress that the label "control group" does not mean new buildings do not impact these cells. My empirical strategy can only identify differential effects between neighborhoods more or less exposed. I choose to use this term to align with the difference-in-differences literature.

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

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. Some of these variables are included in level and variation prior to 2006. Table C1 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 from a treated/control cell. The reason for including the last group will be explained below. Table C2 presents the parameters obtained for the chosen lambda, and Figure C5 exhibits the spatial distribution of fitted values.[26]

My objective is to select outer-ring cells that, on average, have a probability of being treated similar 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 m radius circumference drawn from the centroid of each cell in the initial sample. 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.

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 128 cells (64 first-ring, 64 outer-ring), and the second sample contains 268 cells (134 second-ring, 134 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

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

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 in the initial sample, depicted in Table D3, reveals that the matching significantly enhances the similarity of treated and control groups.[27]

## 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 $\Psi_c$ and an interaction of district ($d$) and time indicators $\mu_{d,t}$.[28]

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 $\alpha_r$ is the treatment

[27] Since the econometric specification controls for district-specific trends, treated cells in districts without any control cell need to be dropped from the estimation for identification. After the matching, I discard three pairs of treated/control cells in the first sample and one pair in the second sample. The location of these cells can be found by inspection of Figure 4.

[28] 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".

effect for group $r$ now averaged in time.

I estimate Equations (13) and (14) separately for each sample using the two-stage procedure proposed by Gardner (2021), which consists of regressing $y_{c,t}$ on $\Psi_c$ 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 and periods. Standard errors are clustered at the cell level.[29]

# 5 Results

I now present estimates of new building effects 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 first analyze the effects of new commercial buildings on the aggregate level of economic activity. Figure 5 displays plots of results from Equation (13) for various outcomes. The upper left and right panels show, respectively, that the number of establishments and workers in first-ring cells are differentially impacted, and the positive effects persist over the period of analysis. Importantly, these panels show no sign of pre-trends prior to the beginning of treatment. In the lower left and right panels of Figure 5, I display the effects of new buildings on the average wage premium and the share of college-educated workers, respectively, two measures potentially correlated with local productivity. They show no evidence that treated

[29] 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 to my setting is that it allows me to account for specific trends that might confound the results. See de Chaisemartin and D'Haultfoeuille (2022) for a detailed discussion.

cells were affected.

While the upper panels indicate that the level of economic activity in first-ring cells is differentially affected, the lower panels suggest that local productivity remains stable in treated cells. According to the theoretical discussion in Section 3, we should expect both variables to be affected if the employment shocks associated with the new buildings generate spillovers on productivity. However, it is possible that such effects are concentrated on specific sectors, which 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 the treatment's initiation. Despite not being possible to reject the null hypothesis at the 5% significance level, the results also suggest an increase in the number of high- and low-skilled offices. Figure 7 shows a similar pattern regarding employment, with persistent effects on local services and offices.

The employment growth of these sectors in first-ring cells can be interpreted, through the lens of the model, as a manifestation of the two types of agglomeration forces considered in this paper. The higher presence of offices would indicate that new buildings boost local productivity through spillover effects, whereas the higher presence of local services would indicate an increase in local demand for non-tradable goods.

To further examine this hypothesis, I estimate how wages and the share of college-educated 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 point estimates after the treatment starts that persists over the period of analysis, although it is not possible to reject the null hypothesis at the 5% significance level in some periods. High-skilled offices also experience an increase in the share of college-educated workers in first-ring cells, as displayed in Figure 9. In this case, non-offices seem to be impacted as well.

Figures 8 and 9 reinforce the notion 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 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 not considered in the model. If different sectors employ different types of workers and local services observe a more elastic labor supply curve, this sector may experience a negligible impact on wages.

## 5.2 Standard Static Difference-in-Differences

I now examine the results from Equation (14) on the outcomes considered so far, which are summarized in Table 3. Panel A displays the effects on the aggregate economy. Consistent with the event-study plots, Columns (1) and (2) exhibit impacts of 3.8% and 12.9% in the number of establishments and workers, respectively. Likewise, Columns (3) and (5) show that the share of college-educated workers and wages were unaffected.

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 8.6% increase in establishments and a 19.9% increase in employment in first-ring cells. High-skilled offices also experience an increase of 3 percentage points in the share of college-educated workers and an increase of 5.4% in the average wage premium. For local services, establishments and employment increase by 8.3% and 14.3%, respectively. It is worth noting that the expansion of both sectors in first-ring cells occurs at the intensive and extensive margin, i.e., more establishments and more workers per establishment. However, the growth at the intensive margin seems to be more salient for high-skilled offices.

From the coefficients shown in Table 3, it is possible to derive a multiplier effect of the tradable sector on the non-tradable sector. Considering the baseline values, first-ring cells observe an increase of about 47 workers in local services and 92 workers in high-skilled offices, meaning that roughly one job is created in local services for every two additional jobs in high-skilled offices. This number is three times lower than the one obtained by Moretti (2010), 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 interconnections are symmetric, except when considering new buildings' direct effect on local demand. Given that new buildings are mostly occupied by offices, there is a large demand shock for local services that might be partially supplied by other neighborhoods, thus resulting in a potential overestimation of the multiplier effect (recall that neighborhoods that receive a new building are dropped from the estimation).

I analyze this possibility more carefully in Section 6.4 by estimating the local effects of new residential buildings. Since there are no reasons to think these constructions would affect local productivity, tradable sectors are unlikely to expand their presence in treated neighborhoods, and an eventual impact on local services would be more easily interpreted as a direct local demand effect. Table D6 shows weak evidence of more economic activity in treated cells and, in particular, weak evidence that local services are affected.

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

## 5.3 Changes in Firm Composition Within Sectors

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

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

This evidence suggests that changes in firm composition are relevant to understanding why wages increase in high-skilled offices. Two effects might be at play, or a combination of both. One possibility is a higher inflow of highly productive firms that rely more heavily on college-educated workers and pay higher wages. The second possibility is a higher outflow of low-quality firms, perhaps due to increased rental costs.[30]

## 5.4 Floor Space Supply

As the number of establishments differentially increases in first-ring cells (driven mainly by local services), it is worth asking whether the floor space supply responds to this shift in local demand. To check for this possibility, I estimate the effect of new buildings on the stock of commercial floor space computed from IPTU data. Figure 10 shows no evidence of an impact in the period analyzed.[31]

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

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

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

## 5.5 Productivity Spillovers

The results 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 if we consider that the employment impact on high-skilled offices is accompanied by effects on wages and worker composition that do not occur in other sectors.

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

However, it is conceivable that being in a high-employment neighborhood raises the likelihood of closing 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 a 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 productivity spillover hypothesis, they are still based on externalities enhanced by density.

Another possible explanation is that neighborhoods closer to new buildings may become more appealing to workers. For example, if new buildings trigger improvements in local amenities or if their location is closely connected with improvements in labor market access (e.g., public investments in infrastructure).[32] In this case, more firms would choose to locate

[32] 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, which suggests that this channel is of secondary importance. Diamond and McQuade (2019), on the other hand, show evidence of heterogeneous effects depending on the neighborhood income level. They report an increase in house prices in low-income neighborhoods and a decline in high-income neighborhoods.

in these neighborhoods as they would find it easier to recruit workers (Tsivanidis, 2022; Perez et al., 2022). High-skilled offices might be particularly benefited since they are more reliant on skilled workers, which are scarcer. However, in this scenario, wages are expected to decrease according to the model, a prediction that contrasts with the results.[33]

# 6 Robustness Checks

This section summarizes a series of tests to validate the robustness of the findings reported in this paper. Appendix D provides more details. Some specific exercises show a lower employment impact on local services (which might suggest an overestimation of the multiplier effect) and a lower wage impact on high-skilled offices, but overall, the results remain consistent. I also perform an estimation of the effects of new residential buildings to present further evidence consistent with my interpretation of the results.

## 6.1 Larger Sample of Neighborhoods

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

Departing again from the cells within 1 km of a new building site, I now select all cells with at least one office worker in all periods (high- or low-skilled). I identify treated and control groups and perform the matching similarly using the new sample. Figure D1 illustrates this procedure, and Table D1 presents the results from Equation (14), which replicates the structure of Table 3. Note that, in this case, the samples used in each estimation differ because some cells may not contain all sectors.[34]

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

[34]Note also that, in this case, the number of observations in Column (1) is higher because the analysis

Some differences between the two sets of results are worth mentioning. First, Table D1 reports considerably smaller effects on local services (establishments and workers). Although the elasticities suggest that the multiplier effect is smaller than what is reported in Section 5, recall that the samples used in each estimation are different, which prevents an accurate evaluation of the relative growth across sectors. Secondly, now the effects on establishments and workers for non-offices are substantial.

## 6.2 Alternative Thresholds - New Buildings

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

Table D2 displays estimates of $\alpha_r$ from Equation (14) for different thresholds indicated in the columns, focusing exclusively on first-ring cells. Column (1) replicates the baseline results when the average employment threshold $n$ is 500 and the average share of college-degree workers $sh$ is 25%. In Columns (2) to (5), I consider different values of $n$ and $sh$. Overall, the results do not change significantly. In all cases, there is an increase in economic activity driven by local services and high-skilled offices, although the effects on establishments tend to be slightly smaller. High-skilled offices also experience increases in wages and the share of college-degree workers.

## 6.3 Estimation with 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 478 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.

includes establishments with zero employment records.

Table D3 presents a balance test of this sample. Columns (1) to (3) display pre-treatment 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. Compared to Table 2, the differences between treated and control groups are more pronounced. Outer-ring cells exhibit considerably less economic activity, and their firms pay lower wages and are less intensive in skilled labor.

Table D4 shows estimates of Equation (14) using this new sample. In general, the point estimates are lower than those from Table 3. Notably, the effects on the wages and the share of college graduates of high-skilled offices are of lower magnitude and statistical significance. The multiplier effect in this case is about 0.4, somewhat lower than what I obtain from Table 3.

In summary, this exercise shows that although the matching substantially improves the similarity between treated and control groups, the estimates do not change dramatically, except perhaps for the wage and worker composition effects on high-skilled offices.

### 6.4 Effects of Residential Buildings

A concern of a different nature is whether the results I report indeed stem from commercial buildings. One plausible hypothesis is that new commercial buildings correlate with surges in residential density. Consequently, offices and local services may choose to locate nearby to benefit from proximity to labor supply or greater local demand.

A possible way to investigate this issue is to estimate the effects of large residential buildings on economic activity. Assessing how closely these findings align with those in Table 3 would shed some light on this matter. Moreover, it can also improve our understanding of the underlying channels through which commercial buildings may affect the local economy. Since residential buildings are unlikely to increase the productivity of nearby businesses, they could provide a sense of whether buildings affect local demand directly.

To do so, 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 shock. Table D5 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).[35]

Then, I repeat the procedure described in Section 4.2 to select the neighborhoods for the analysis and to construct the treatment variables. Figure D2 shows the location of these buildings, together with treatment and control cells. Compared to the sample of commercial buildings, they are more spread throughout the city and farther from high employment areas. This fact *per se* already suggests that large residential and commercial buildings do not have a close connection. It is also worth noting that fewer treated and control cells are selected for the analysis as most cells within 1 km do not contain firms from all sectors analyzed.

Table D6 displays the results from Equation (14) in this new setting for the same sectoral variables examined so far.[36] 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. In particular, there is weak evidence that local services are affected somehow.[37]

Hence, the results show that the opening of commercial and residential buildings and their local effects are weakly correlated. Additionally, the lack of impact on local services reinforces the idea that commercial buildings directly affect the local economy through productivity spillovers, and the higher presence of local services occurs due to increased economic activity within the neighborhoods.

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

[36] In this case, I do not perform any matching procedure as the lasso estimation returns a trivial solution with all coefficients equal to zero, except for the intercept. In other words, no linear combination of the regressors is helpful in predicting the construction of large residential buildings.

[37] The results are qualitatively similar if I impose more flexible criteria to select the cells, as in Section 6.1.

## 6.5 Other tests

Appendix D also provides results using continuous treatment based on the distance to new buildings (Table D7) and an alternative clustering of standard errors based on the closest new building (Table D8). Overall, the conclusions remain unchanged, except that the estimated impact on local services' employment displayed in Table D7 is smaller.

Finally, I also report in Figures D3-D6 event-study plots using the doubly-robust estimator proposed by Callaway and Sant'Anna (2021). In this case, I use all outer-ring cells as controls and the PPS as the covariate. The figures exhibit qualitatively the same patterns observed in Figures 6-9, except that now there is some sign of pre-trends in the share of college graduates in high-skilled offices. Moreover, confidence intervals tend to be larger in general.

I consider that Callaway and Sant'Anna (2021)'s estimator is less suitable than Gardner (2021)'s in this setting for two reasons. First, for some years, I have only a small number of observations to compute the group-time ATTs necessary in Callaway and Sant'Anna (2021), which makes the estimation more noisy. Secondly, Gardner (2021) gives me more flexibility to control for distric-specific trends.

# 7 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 estimation 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 two 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 influenced to some degree by the sorting of high-wage firms in treated neighborhoods.

My findings corroborate 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[38].

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

[38]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).

# 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.

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

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

Gardner, J. (2021). 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.

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.

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

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

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

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

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

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

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

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

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 Finance*, 140:106510.

# Tables

Table 1. Summary Statistics: New Commercial Buildings

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

*Notes:* 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.

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 | 268.4 (383.4) | 326.1 (715.7) | [0.326] | 160.0 (384.1) | 197.5 (532.0) | [0.252] |
| Low-skilled Offices | 176.1 (437.3) | 316.2 (843.5) | [0.042] | 230.7 (731.1) | 223.1 (786.9) | [0.888] |
| Local Services | 260.2 (457.4) | 277.5 (533.9) | [0.735] | 180.5 (250.1) | 214.8 (510.6) | [0.227] |
| Non-offices | 289.5 (358.6) | 257.6 (394.3) | [0.409] | 199.7 (299.3) | 175.1 (289.3) | [0.237] |
| Establishments | | | | | | |
| High-skilled Offices | 15.7 (14.9) | 15.0 (14.7) | [0.649] | 10.8 (11.8) | 10.2 (11.6) | [0.462] |
| Low-skilled Offices | 14.3 (17.6) | 13.6 (12.1) | [0.663] | 13.1 (12.8) | 12.6 (13.0) | [0.541] |
| Local Services | 24.8 (37.6) | 26.3 (33.1) | [0.679] | 21.1 (17.1) | 23.9 (41.2) | [0.207] |
| Non-offices | 14.4 (11.2) | 13.2 (10.6) | [0.301] | 12.0 (9.2) | 10.9 (8.6) | [0.070] |
| Wages | | | | | | |
| High-skilled Offices | 5709.4 (4429.2) | 5391.0 (5282.7) | [0.523] | 4056.6 (3373.8) | 4211.6 (4207.4) | [0.565] |
| Low-skilled Offices | 3297.7 (2457.0) | 2787.8 (2435.9) | [0.042] | 2603.4 (2123.9) | 2300.0 (1912.1) | [0.034] |
| Local Services | 2797.4 (2041.8) | 2598.8 (1948.1) | [0.330] | 2467.3 (1642.0) | 2216.6 (1581.9) | [0.028] |
| Non-offices | 5840.3 (4228.7) | 4682.4 (3326.8) | [0.003] | 4016.8 (2822.4) | 4072.4 (3066.4) | [0.789] |
| % College | | | | | | |
| High-skilled Offices | 44.3 (25.3) | 47.5 (21.9) | [0.184] | 39.4 (21.1) | 40.4 (23.0) | [0.543] |
| Low-skilled Offices | 29.1 (19.7) | 23.6 (17.3) | [0.004] | 23.7 (18.9) | 20.2 (16.1) | [0.005] |
| Local Services | 18.2 (15.5) | 16.9 (15.5) | [0.404] | 15.6 (14.6) | 13.5 (14.1) | [0.034] |
| Non-offices | 33.6 (21.5) | 32.2 (20.3) | [0.530] | 29.4 (19.8) | 29.5 (19.7) | [0.925] |
| Observations | 64 | 64 | | 134 | 134 | |

*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.

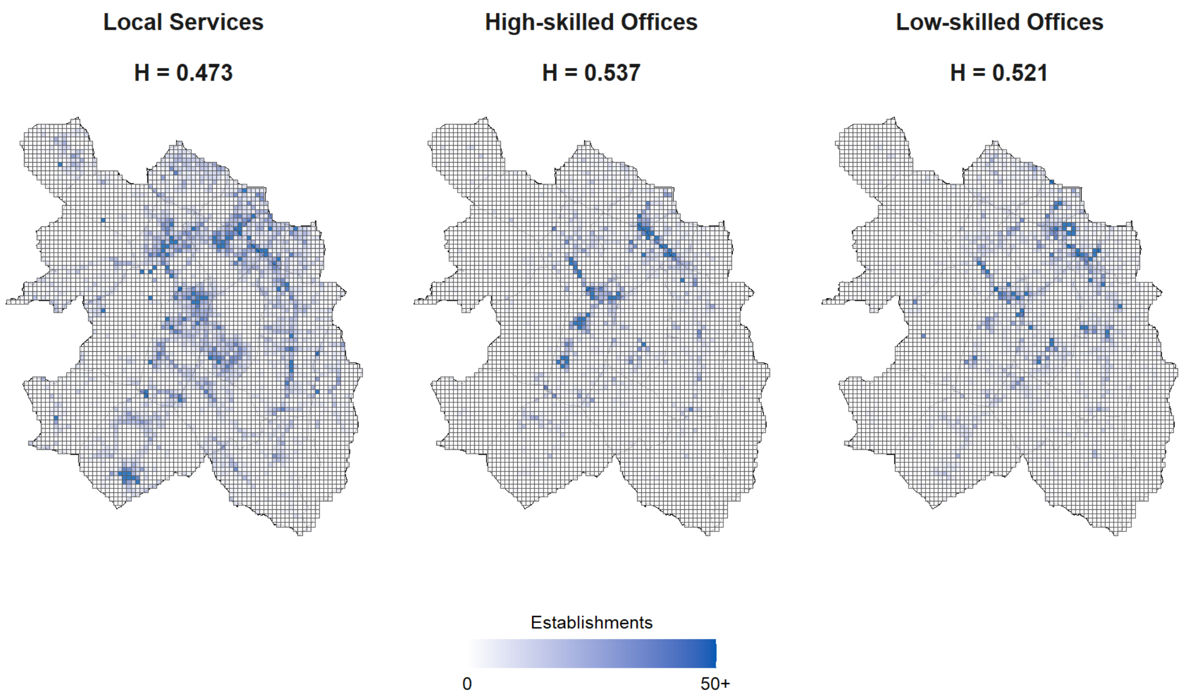
Table 3. Effects of New Commercial Buildings: DiD Results

| | Log Estabs | Log Workers | % College | | Wage Premium | |
|---|---|---|---|---|---|---|
| | | | All estabs | excl. new estabs | All estabs | excl. new estabs |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| **Panel A. All Sectors** | | | | | | |
| 0-250m | 0.0378 | 0.1288** | 0.0060 | | 0.0073 | |
| | (0.0252) | (0.0536) | (0.0108) | | (0.0171) | |
| $R^2$ | 0.00830 | 0.02090 | 0.00070 | | 0.00017 | |
| Observations | 1,920 | 1,920 | 1,920 | | 1,920 | |
| 250-500m | 0.0082 | 0.0024 | -0.0020 | | -0.0025 | |
| | (0.0141) | (0.0396) | (0.0064) | | (0.0104) | |
| $R^2$ | 0.00035 | -0.00024 | -0.000008 | | -0.00016 | |
| Observations | 4,020 | 4,020 | 4,020 | | 4,020 | |
| **Panel B. High-Skilled Offices** | | | | | | |
| 0-250m | 0.0860* | 0.1994** | 0.0296** | 0.0216 | 0.0539** | 0.0150 |
| | (0.0476) | (0.0878) | (0.0150) | (0.0141) | (0.0272) | (0.0224) |
| $R^2$ | 0.01148 | 0.01668 | 0.01293 | 0.00799 | 0.01351 | 0.00110 |
| Observations | 1,920 | 1,920 | 1,920 | 1,830 | 1,920 | 1,830 |
| 250-500m | 0.0168 | 0.0552 | 0.0014 | -0.0073 | 0.0129 | -0.0054 |
| | (0.0294) | (0.0699) | (0.0102) | (0.0112) | (0.0170) | (0.0170) |
| $R^2$ | 0.00022 | 0.00078 | -0.00022 | 0.00041 | 0.00061 | -0.00011 |
| Observations | 4,020 | 4,020 | 4,020 | 3,645 | 4,020 | 3,645 |
| **Panel C. Low-Skilled Offices** | | | | | | |
| 0-250m | 0.0747* | 0.1481 | -0.0016 | -0.0214 | -0.0256 | -0.0380 |
| | (0.0424) | (0.1227) | (0.0136) | (0.0132) | (0.0232) | (0.0233) |
| $R^2$ | 0.01040 | 0.00483 | -0.00049 | 0.00607 | 0.00326 | 0.00878 |
| Observations | 1,920 | 1,920 | 1,920 | 1,770 | 1,920 | 1,770 |
| 250-500m | 0.0414 | 0.1038 | -0.0126 | -0.0218** | -0.0090 | -0.0545*** |
| | (0.0275) | (0.0769) | (0.0106) | (0.0106) | (0.0145) | (0.0140) |
| $R^2$ | 0.00333 | 0.00266 | 0.00191 | 0.00657 | 0.00026 | 0.01952 |
| Observations | 4,020 | 4,020 | 4,020 | 3,780 | 4,020 | 3,780 |
| **Panel D. Local Services** | | | | | | |
| 0-250m | 0.0833** | 0.1434** | -0.000005 | -0.0107 | -0.0074 | -0.0325** |
| | (0.0401) | (0.0648) | (0.0085) | (0.0090) | (0.0138) | (0.0158) |
| $R^2$ | 0.01516 | 0.01982 | -0.00052 | 0.00353 | 0.00040 | 0.01181 |
| Observations | 1,920 | 1,920 | 1,920 | 1,905 | 1,920 | 1,905 |
| 250-500m | 0.0139 | -0.0019 | -0.0097* | -0.0117** | -0.0163* | -0.0223** |
| | (0.0180) | (0.0339) | (0.0056) | (0.0055) | (0.0092) | (0.0094) |
| $R^2$ | 0.00060 | -0.00024 | 0.00381 | 0.00628 | 0.00382 | 0.00737 |
| Observations | 4,020 | 4,020 | 4,020 | 3,945 | 4,020 | 3,945 |
| **Panel E. Non-Offices** | | | | | | |
| 0-250m | -0.0170 | 0.0967 | 0.0270* | 0.0260* | 0.0102 | 0.0004 |
| | (0.0565) | (0.0906) | (0.0145) | (0.0155) | (0.0228) | (0.0210) |
| $R^2$ | 0.00004 | 0.00451 | 0.01278 | 0.01088 | 0.00017 | -0.00054 |
| Observations | 1,920 | 1,920 | 1,920 | 1,845 | 1,920 | 1,845 |
| 250-500m | -0.0341 | -0.0715 | 0.0049 | 0.0130 | -0.0041 | -0.0311* |
| | (0.0265) | (0.0524) | (0.0093) | (0.0107) | (0.0142) | (0.0160) |
| $R^2$ | 0.00248 | 0.00251 | 0.00013 | 0.00211 | -0.00014 | 0.00534 |
| Observations | 4,020 | 4,020 | 4,020 | 3,870 | 4,020 | 3,870 |

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

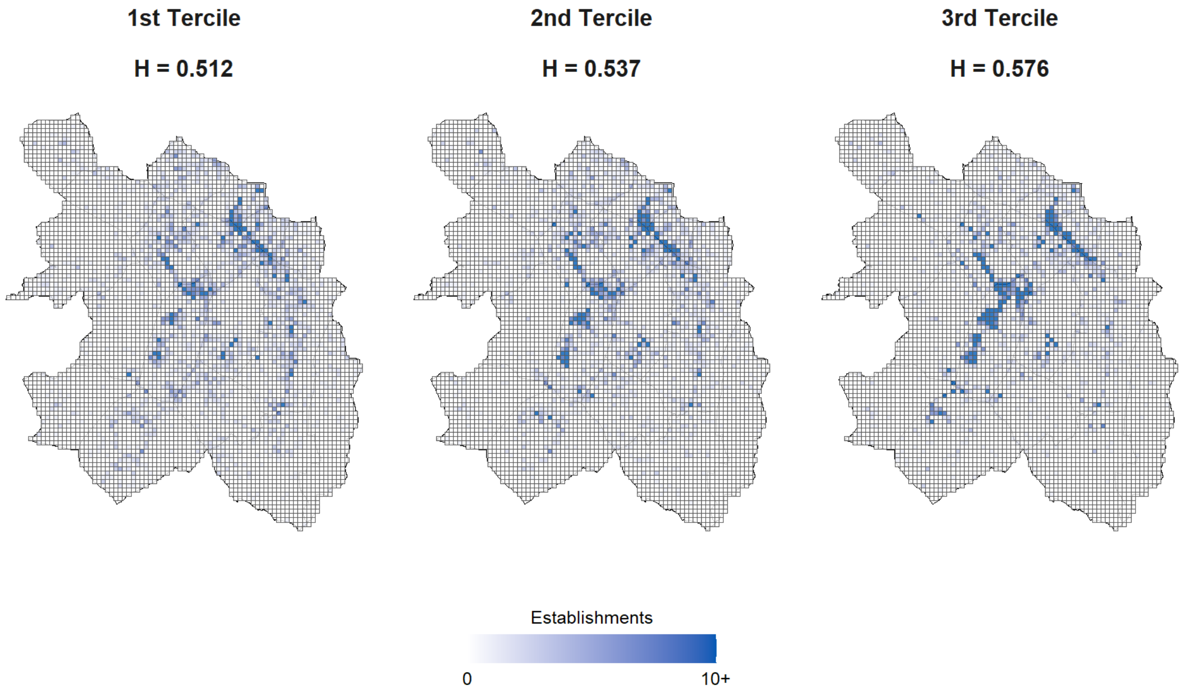
# Figures

Figure 1. Spatial Distribution of Employment by Sector



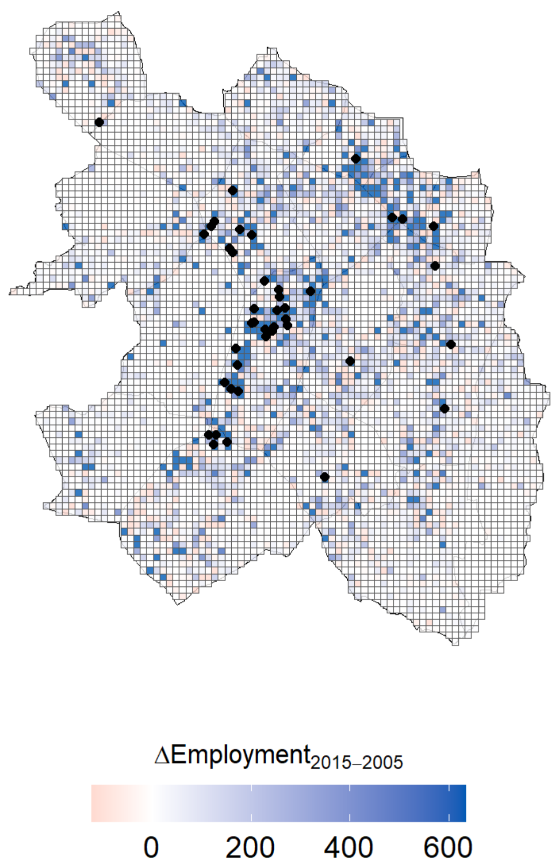

*Notes:* This figure displays, for 2010 , the spatial distribution of establishments by cell for different sectors.

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



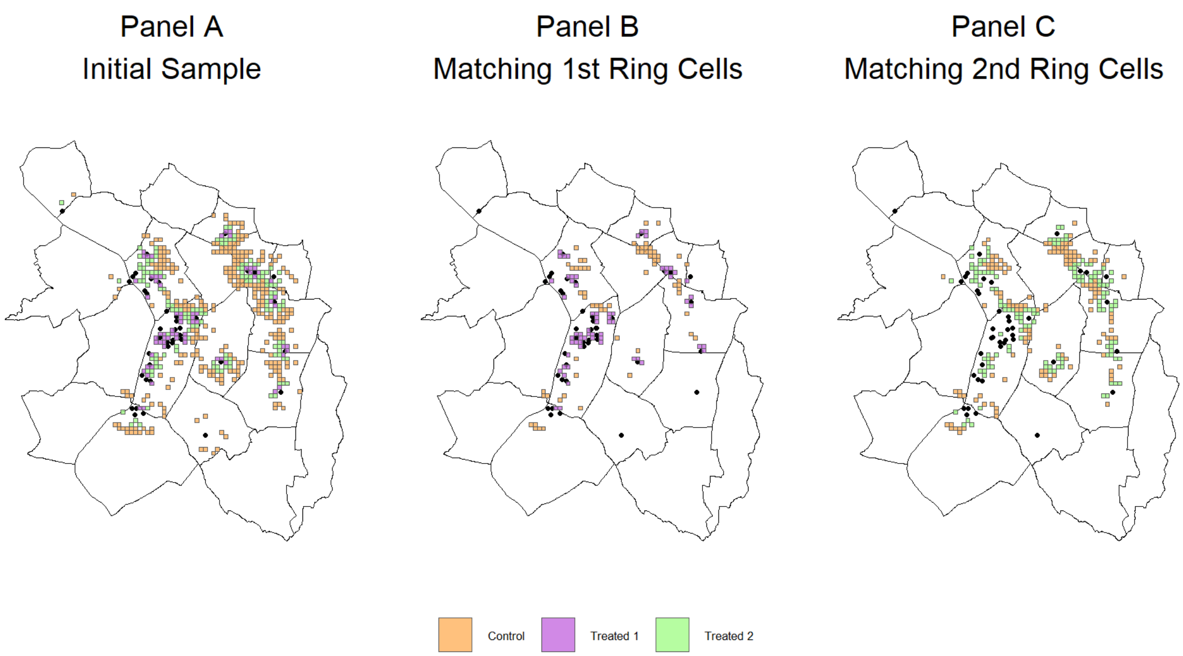

*Notes:* This figure displays, for 2010, the spatial distribution of high-skilled offices by cell for different terciles of establishment wage premium. For more details about the estimation of these premia, see Section 2.1.

Figure 3. Location of New Commercial Buildings in São Paulo



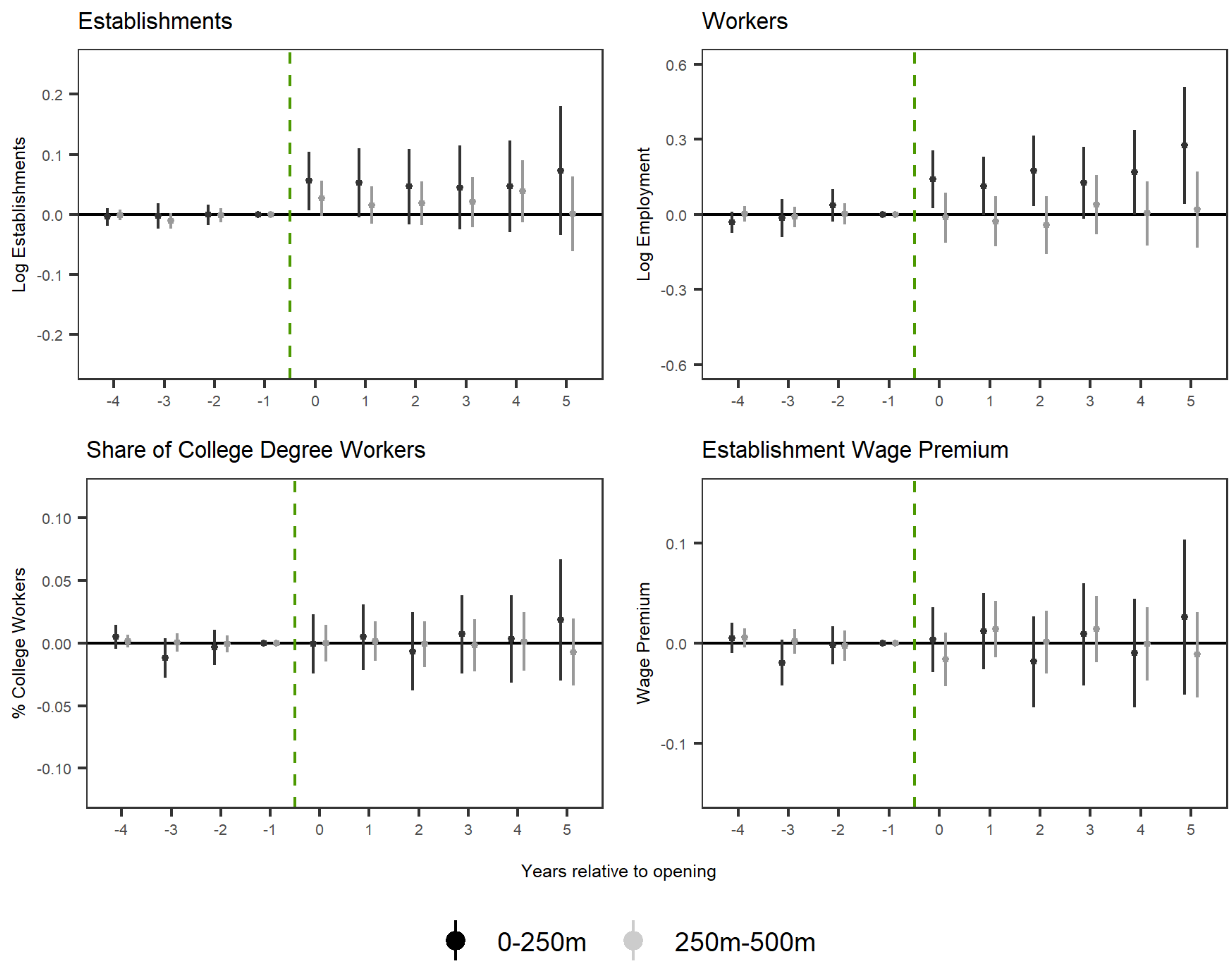

*Notes:* This figure shows in black dots the location of the new commercial buildings considered in the analysis, together with cell variation in employment between 2005 and 2015.

Figure 4. Empirical Analysis Setup



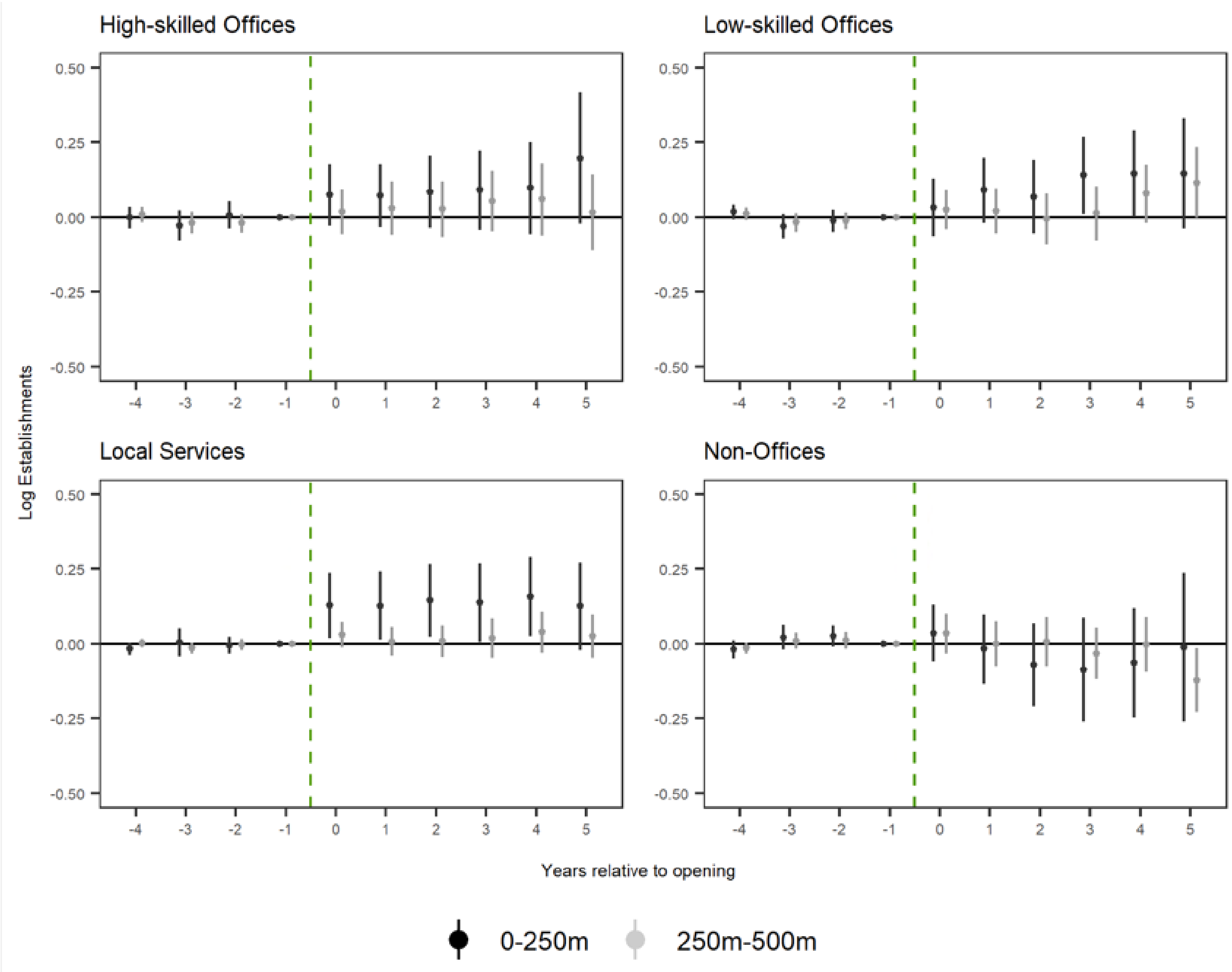

*Note:* This figure depicts the design of the empirical analysis. Panel A shows the initial sample of treated and control cells. The black dots represent the new commercial buildings, and the solid lines represent district borders of São Paulo. Panels B and C present the results of the matching procedure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

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



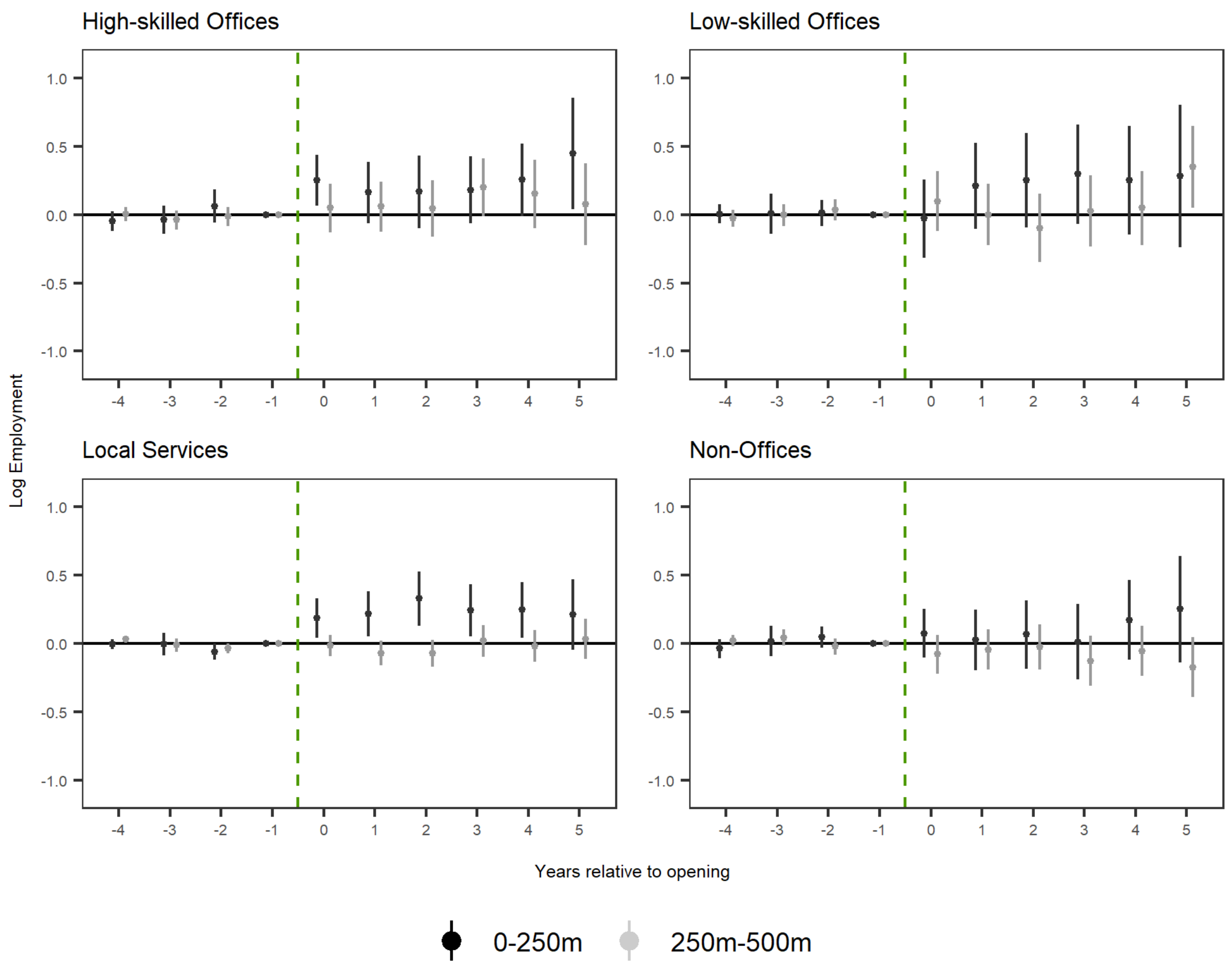

*Note:* This figure plots coefficients from running Equation (13) on the log number of establishments (upper left panel), the log employment (upper right panel), the share of workers with college degree (lower left panel) and the wage premium (lower right panel). The bars indicates the 95% confidence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

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



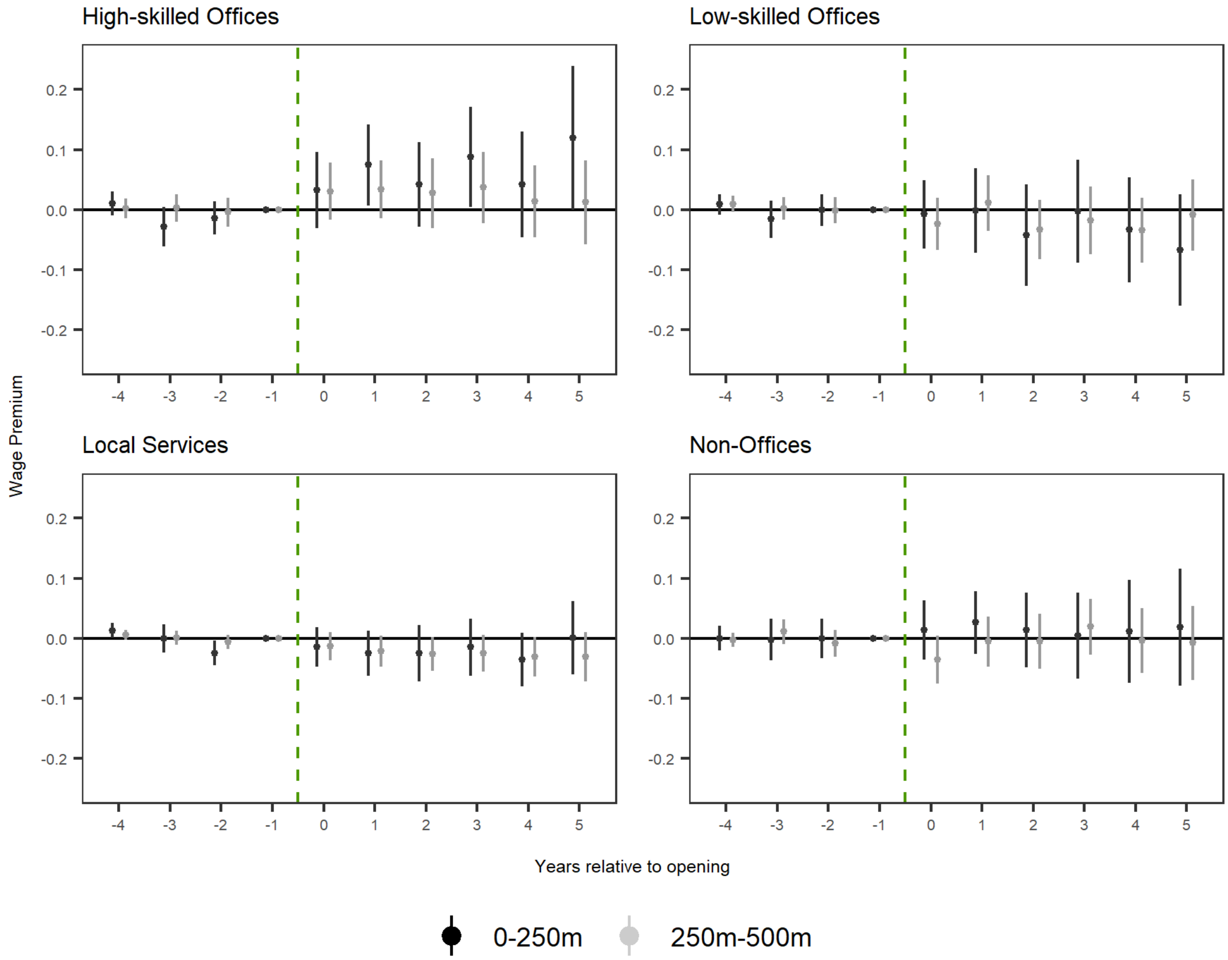

*Note:* This figure plots coefficients from running Equation (13) on the log number of establishments for different sectors. The definition of each sector is described in Section 2.2. The bars indicates the 95% confidence interval, where standard errors are clustered at the cell level.

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



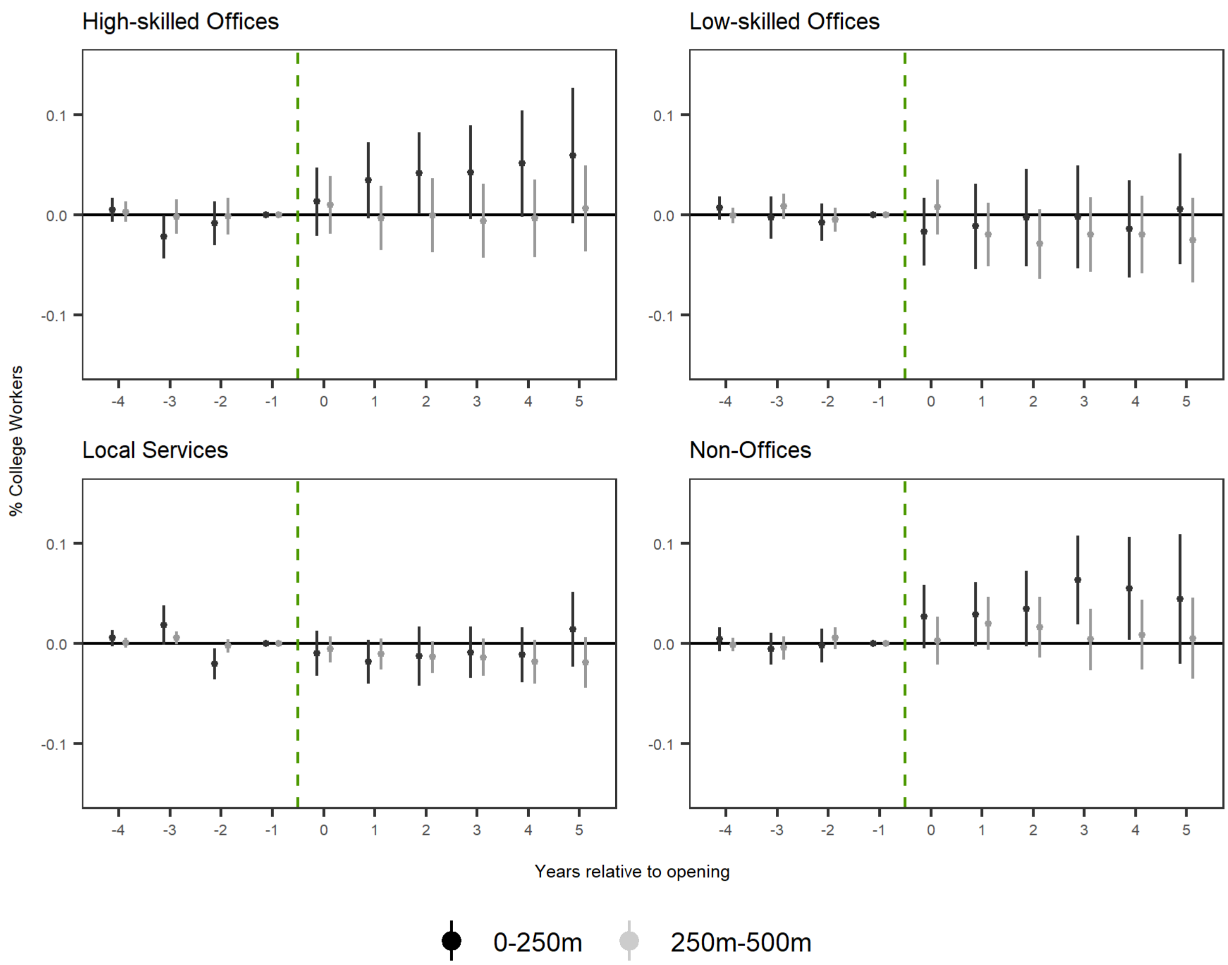

*Note:* This figure plots coefficients from running Equation (13) on log employment for different sectors. The definition of each sector is described in Section 2.2. The bars indicates the 95% confidence interval, where standard errors are clustered at the cell level.

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



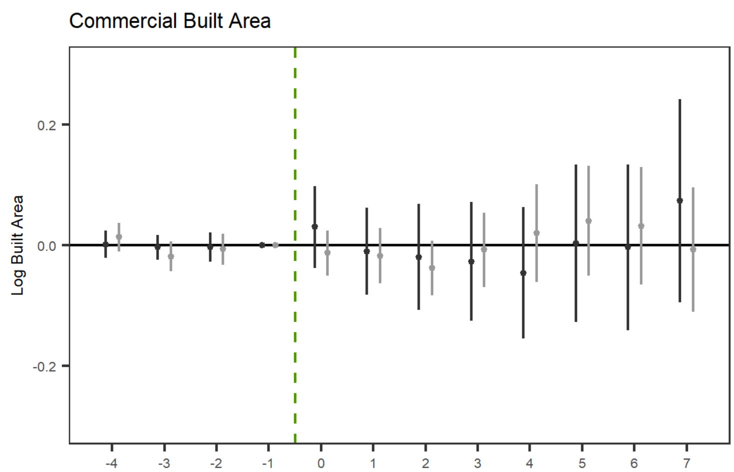

*Note:* This figure plots coefficients from running Equation (13) on mean establishment wage premium (weighted by establishment size) for different sectors. The definition of each sector is described in Section 2.2. The bars indicates the 95% confidence interval, where standard errors are clustered at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

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.

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.

# Appendix

## Contents

# A - 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.

## B - Theoretical Results and Derivations

### B.1 Model Derivations

*Firm Labor Supply.*— Consider individuals $i$ living in different neighborhoods $m$ who need to choose a firm e to work. They take into account where the firm is located for two reasons. First, they spend a fraction $\delta$ of their wages on local services, whose price is neighborhood-specific. Secondly, some neighborhoods offer higher utility than others.[39]

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Combining (B.9) and (7) yields

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

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

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

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

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

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

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

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

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

Finally, I rearrange the terms to get

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

## B.2 Proof of Proposition 1

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

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

and

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

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

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

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 (5). Now, combining expressions (5) and (B.13) and taking the derivative of $E_{s,n}$ with respect to $A_{s,n}$:

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

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

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

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

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

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

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

### B.3 Proof of Proposition 2

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

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

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

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

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

For local services, the same procedure yields

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

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

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

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

and

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

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

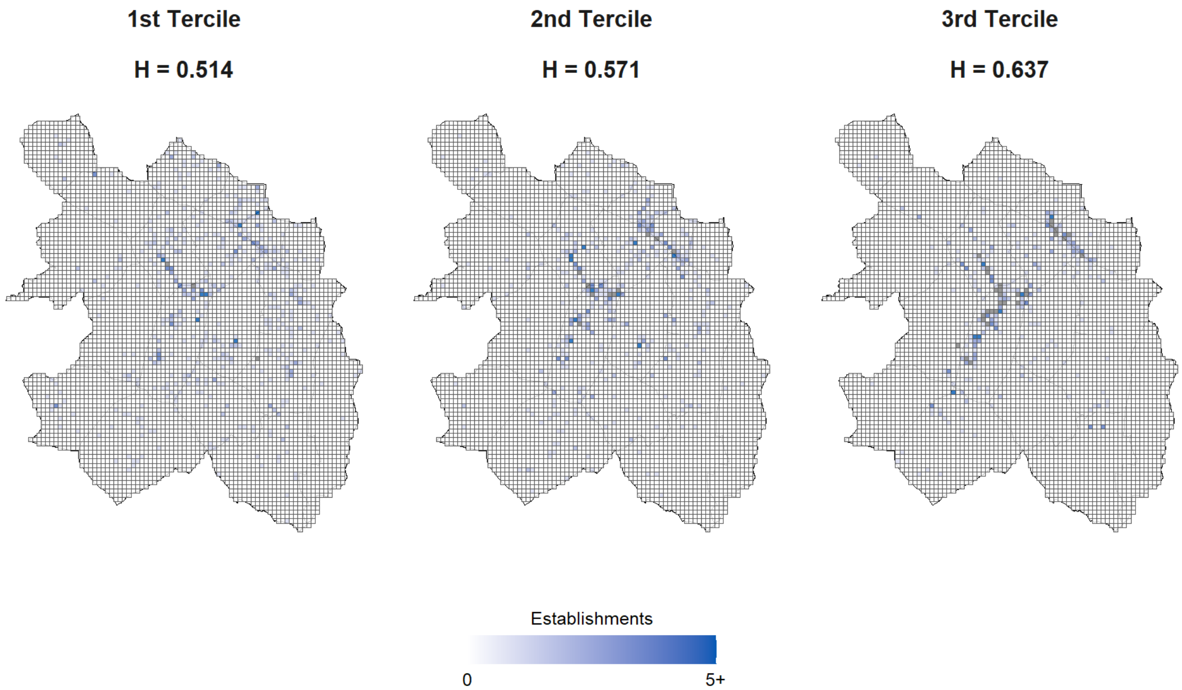
$$\begin{aligned}
\frac{\partial w^R_{s,n}}{\partial A_{s,n}} &= \frac{M}{1-\beta} p_s^{\frac{1}{1-\beta}} \left(\frac{A_{s.n}}{p_{LS,n}^{\beta\kappa+\delta(1-\beta)}}\right)^{\frac{1}{1-\beta}-1} \left[\frac{1}{p_{LS,n}^{\beta\kappa+\delta(1-\beta)}} - (\beta\kappa+\delta(1-\beta))\frac{A_{s.n}}{p_{LS,n}^{\beta\kappa+\delta(1-\beta)+1}}\frac{\partial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

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 (B.18) shows that the derivative is positive if $\beta\kappa + \delta(1-\beta) < 1$, which is not guaranteed as well.

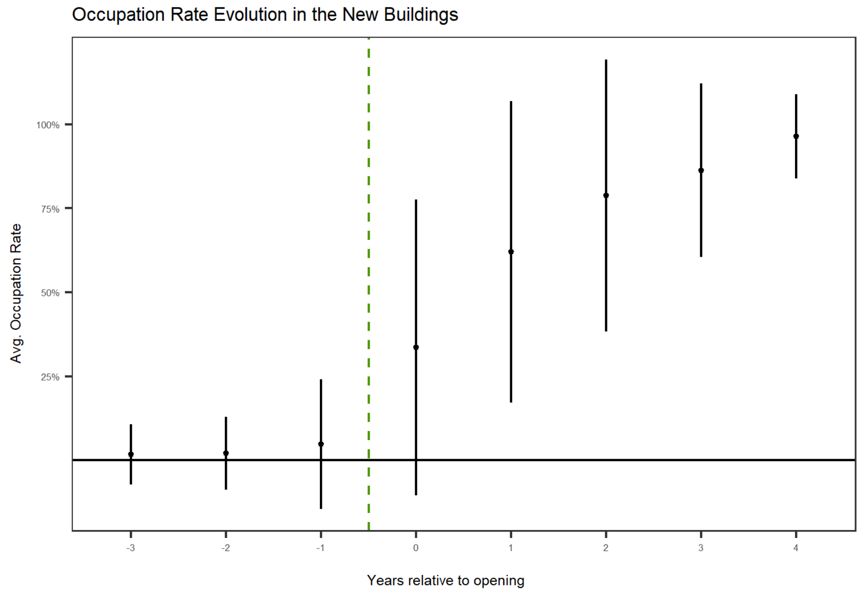
# C - Additional Figures and Tables

Figure C1. Spatial Distribution of Finance Establishments




*Notes:* 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 C2. 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.

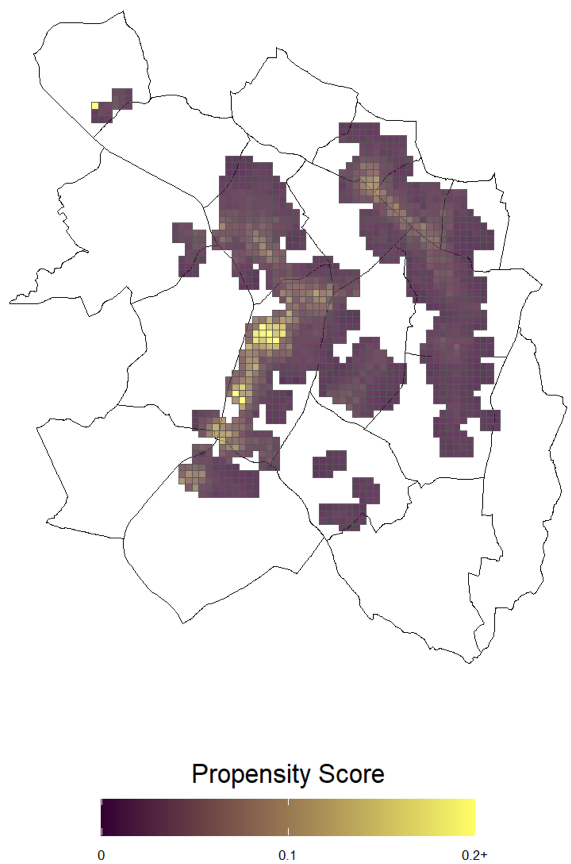
Figure C3. Examples of Commercial Buildings


*Source:* Google Maps.

Figure C4. Validating Selected Buildings


*Source:* Google Maps. *Notes:* Example of a commercial building inaugurated in 2013. Using Google Maps imagery from 2011 and 2014, it is possible to check if the timeline of construction is consistent with the date of inauguration defined in Section 4.

Figure C5. Fitted values: Propensity Score




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

Table C1. List of Variables used in the Propensity Score Model

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

Table C2. Coefficients: Propensity Score Model

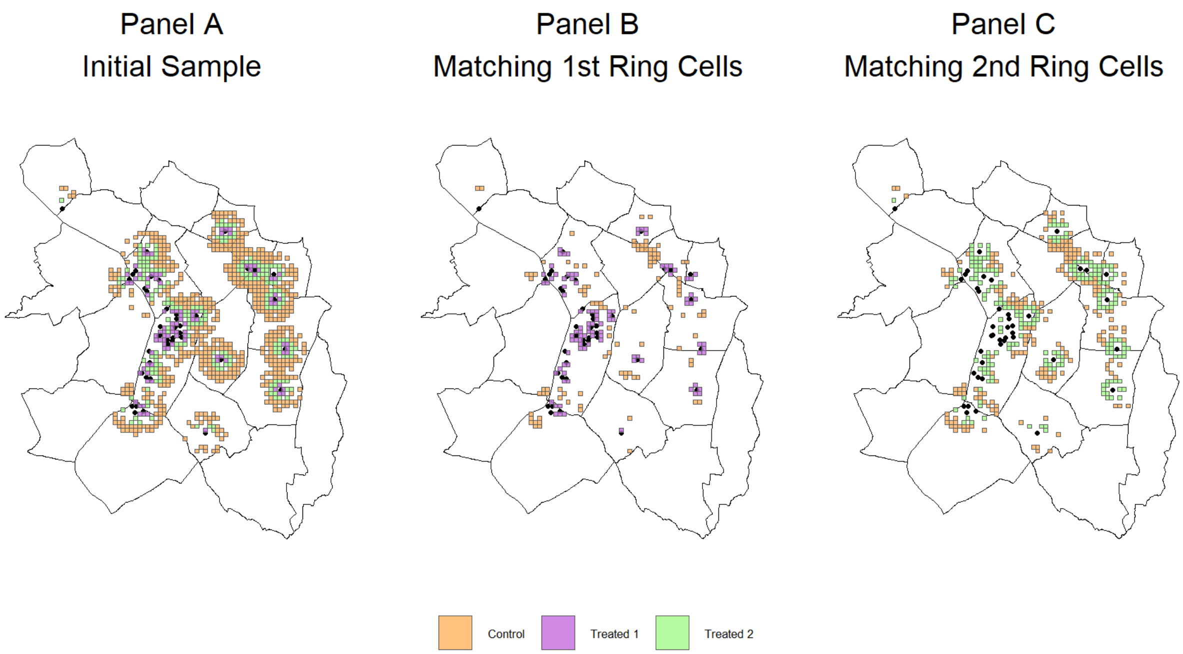
| Variable | Estimate |
|---|---|
| Log built area growth commercial – cell | 0.360 |
| Log (1+ employment adm. and support) – cell | 0.007 |
| Log employment growth finance – cell | 0.055 |
| Log (1+ employment finance) – cell | 0.025 |
| Log (1+ employment information and communication) – buffer |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

*Notes:* This table shows the estimated coefficients of Equation (12) using Lasso. For a description of all variables used, see Table C1

# D - Robustness Checks

## Larger Sample of Neighborhoods

Figure D1. Empirical Analysis Setup




*Notes:* This figure depicts the design of the empirical analysis using an alternative sample of cells, as described in Section 6.1. 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 D1. Effects of New Commercial Buildings: Larger Sample of Neighborhoods

| | Log Estabs | Log Workers | % College | | Wage Premium | |
|---|---|---|---|---|---|---|
| | | | All estabs | excl. new estabs | All estabs | excl. new estabs |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| **Panel A. All Sectors** | | | | | | |
| 0-250m | 0.0481** | 0.1051** | 0.0116 | | 0.0071 | |
| | (0.0207) | (0.0480) | (0.0100) | | (0.0141) | |
| $R^2$ | 0.01031 | 0.01038 | 0.00402 | | 0.00024 | |
| Observations | 3,000 | 3,000 | 3,000 | | 3,000 | |
| 250-500m | 0.0137 | 0.0896*** | 0.0004 | | 0.0044 | |
| | (0.0153) | (0.0338) | (0.0052) | | (0.0089) | |
| $R^2$ | 0.00077 | 0.00650 | -0.00015 | | 0.00009 | |
| Observations | 6,300 | 6,300 | 6,300 | | 6,300 | |
| **Panel B. High-Skilled Offices** | | | | | | |
| 0-250m | 0.1151*** | 0.2159*** | 0.0254* | 0.0103 | 0.0540** | 0.0159 |
| | (0.0436) | (0.0830) | (0.0143) | (0.0138) | (0.0257) | (0.0229) |
| $R^2$ | 0.01610 | 0.01740 | 0.00728 | 0.00109 | 0.01272 | 0.00119 |
| Observations | 2,565 | 2,505 | 2,505 | 2,310 | 2,505 | 2,310 |
| 250-500m | 0.0292 | 0.1739*** | 0.0023 | -0.0079 | 0.0059 | -0.0123 |
| | (0.0304) | (0.0651) | (0.0103) | (0.0107) | (0.0163) | (0.0161) |
| $R^2$ | 0.00096 | 0.00932 | -0.00015 | 0.00052 | 0.00004 | 0.00060 |
| Observations | 5,175 | 4,920 | 4,920 | 4,425 | 4,920 | 4,425 |
| **Panel C. Low-Skilled Offices** | | | | | | |
| 0-250m | 0.0724* | 0.1538 | 0.0076 | -0.0062 | -0.0364* | -0.0464** |
| | (0.0388) | (0.1267) | (0.0116) | (0.0115) | (0.0193) | (0.0197) |
| $R^2$ | 0.00945 | 0.00521 | 0.00045 | 0.00016 | 0.00691 | 0.01213 |
| Observations | 2,640 | 2,610 | 2,610 | 2,415 | 2,610 | 2,415 |
| 250-500m | 0.0347 | 0.0390 | -0.0014 | -0.0107 | 0.0053 | -0.0312** |
| | (0.0249) | (0.0702) | (0.0088) | (0.0093) | (0.0121) | (0.0122) |
| $R^2$ | 0.00209 | 0.00021 | -0.00015 | 0.00144 | -0.000006 | 0.00611 |
| Observations | 5,730 | 5,685 | 5,685 | 5,310 | 5,685 | 5,310 |
| **Panel D. Local Services** | | | | | | |
| 0-250m | 0.0220 | 0.0256 | -0.0002 | -0.0140** | 0.0050 | -0.0283** |
| | (0.0328) | (0.0612) | (0.0074) | (0.0071) | (0.0132) | (0.0118) |
| $R^2$ | 0.00083 | 0.00026 | -0.00035 | 0.00665 | -0.00002 | 0.00896 |
| Observations | 2,880 | 2,865 | 2,865 | 2,850 | 2,865 | 2,850 |
| 250-500m | -0.0246 | -0.0200 | -0.0059 | -0.0066 | -0.0059 | -0.0156* |
| | (0.0193) | (0.0351) | (0.0046) | (0.0046) | (0.0084) | (0.0083) |
| $R^2$ | 0.00206 | 0.00029 | 0.00146 | 0.00196 | 0.00035 | 0.00362 |
| Observations | 5,745 | 5,685 | 5,685 | 5,520 | 5,685 | 5,520 |
| **Panel E. Non-Offices** | | | | | | |
| 0-250m | 0.0671* | 0.1320* | 0.0320*** | 0.0323*** | 0.0114 | -0.0041 |
| | (0.0382) | (0.0716) | (0.0115) | (0.0119) | (0.0197) | (0.0200) |
| $R^2$ | 0.00785 | 0.00797 | 0.01744 | 0.01612 | 0.00040 | -0.00030 |
| Observations | 2,760 | 2,730 | 2,730 | 2,610 | 2,730 | 2,610 |
| 250-500m | -0.0162 | -0.0029 | -0.0031 | 0.0020 | -0.0018 | -0.0233* |
| | (0.0216) | (0.0483) | (0.0076) | (0.0086) | (0.0117) | (0.0130) |
| $R^2$ | 0.00041 | -0.00017 | -0.00002 | -0.00012 | -0.00015 | 0.00315 |
| Observations | 5,775 | 5,700 | 5,700 | 5,475 | 5,700 | 5,475 |

*Notes:* This table reports estimates of $\alpha_r$ in Equation (14) for different outcome variables using alternative samples of cells, as described in Section 6.1. Standard errors clustered at the cell level are displayed in parentheses. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

## Alternative Thresholds - New Buildings

Table D2. Effects of New Commercial Buildings: First-ring Cells

| Variable | Baseline | $n = 600$ | $n = 400$ | $sh = 15\%$ | $sh = 35\%$ |
|---|---|---|---|---|---|
| **High-skilled Offices** | | | | | |
| Log Estabs | 0.086 (0.0476) | 0.0913 (0.0616) | 0.0728 (0.0407) | 0.0851 (0.04) | 0.0101 (0.0514) |
| Log Workers | 0.1994 (0.0878) | 0.3433 (0.1296) | 0.2254 (0.0865) | 0.1662 (0.0959) | 0.1582 (0.116) |
| % College | 0.0296 (0.015) | 0.0606 (0.0202) | 0.0085 (0.0142) | 0.0244 (0.0132) | 0.0329 (0.0181) |
| Wage Premium | 0.0539 (0.0272) | 0.0513 (0.0411) | 0.0613 (0.026) | 0.0632 (0.0274) | 0.0743 (0.0333) |
| **Low-skilled Offices** | | | | | |
| Log Estabs | 0.0747 (0.0424) | 0.1225 (0.0541) | 0.0627 (0.0374) | 0.0653 (0.0356) | 0.0505 (0.0406) |
| Log Workers | 0.1481 (0.1227) | 0.0897 (0.1516) | 0.0619 (0.11) | 0.1405 (0.113) | 0.023 (0.1235) |
| % College | -0.0016 (0.0136) | 0.0091 (0.0185) | -0.002 (0.0126) | -0.0045 (0.0124) | 0.0044 (0.0143) |
| Wage Premium | -0.0256 (0.0231) | -0.0391 (0.0283) | -0.0381 (0.0205) | -0.0368 (0.02) | -0.0252 (0.0245) |
| **Local Services** | | | | | |
| Log Estabs | 0.0833 (0.0401) | 0.1338 (0.055) | 0.039 (0.0398) | 0.0529 (0.0348) | 0.0981 (0.0453) |
| Log Workers | 0.1434 (0.0648) | 0.1857 (0.0851) | 0.0967 (0.0592) | 0.0976 (0.0552) | 0.1742 (0.0717) |
| % College | 0 (0.0085) | 0.0039 (0.0108) | 0.0021 (0.0072) | -0.0032 (0.0072) | 0.0033 (0.008) |
| Wage Premium | -0.0074 (0.0138) | 0.0075 (0.0171) | -0.0027 (0.0129) | -0.0126 (0.0135) | 0.0046 (0.0154) |
| **Non-offices** | | | | | |
| Log Estabs | -0.017 (0.0565) | 0.0139 (0.0716) | 5e-04 (0.0398) | -0.0028 (0.0385) | 0.0265 (0.0497) |
| Log Workers | 0.0967 (0.0906) | 0.1617 (0.1169) | 0.0559 (0.0737) | -0.0127 (0.0612) | 0.0674 (0.0723) |
| % College | 0.027 (0.0145) | 0.0248 (0.0196) | 0.0322 (0.0115) | 0.031 (0.0123) | 0.0273 (0.0145) |
| Wage Premium | 0.0102 (0.0228) | 0.0196 (0.0306) | 0.0208 (0.0208) | 0.0102 (0.0192) | -0.0101 (0.0246) |
| Obs | 1920 | 1380 | 2400 | 2250 | 1740 |

*Notes:* This table displays estimates of $\alpha_r$ in Equation (14) for various outcomes based on alternative thresholds to select new commercial buildings, as described in Section 6.2. 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

Table D3. Baseline Summary Statistics

| Variable | Treated 1 (T1) | Treated 2 (T2) | Control (C) | t-test T1-C | t-test T2-C |
|---|---|---|---|---|---|
| Workers | | | | | |
| High-skilled Offices | 258.7 (377.5) | 158.9 (382.9) | 153.1 (605.2) | [0.002] | [0.839] |
| Low-skilled Offices | 170.4 (428.4) | 229.6 (728.5) | 131.8 (558.9) | [0.283] | [0.018] |
| Local Services | 254.5 (447.9) | 181.9 (249.7) | 150.5 (371.8) | [0.003] | [0.080] |
| Non-offices | 282.2 (352.6) | 198.6 (298.4) | 132.1 (233.0) | [0.000] | [0.000] |
| Establishments | | | | | |
| High-skilled Offices | 15.5 (14.7) | 10.8 (11.7) | 7.4 (9.1) | [0.000] | [0.000] |
| Low-skilled Offices | 14.1 (17.2) | 13.1 (12.8) | 9.7 (11.1) | [0.001] | [0.000] |
| Local Services | 24.9 (36.8) | 21.0 (17.2) | 18.7 (30.7) | [0.029] | [0.104] |
| Non-offices | 14.1 (11.0) | 12.1 (9.2) | 8.8 (6.9) | [0.000] | [0.000] |
| Wages | | | | | |
| High-skilled Offices | 5545.9 (4395.3) | 4046.5 (3363.3) | 3412.2 (3222.2) | [0.000] | [0.002] |
| Low-skilled Offices | 3212.7 (2433.9) | 2597.9 (2117.3) | 2125.1 (1611.9) | [0.000] | [0.000] |
| Local Services | 2790.2 (2011.4) | 2460.7 (1637.8) | 1964.8 (1246.4) | [0.000] | [0.000] |
| Non-offices | 5664.6 (4212.0) | 4015.5 (2812.1) | 3582.2 (2755.0) | [0.000] | [0.011] |
| % College | | | | | |
| High-skilled Offices | 44.5 (24.9) | 39.5 (21.1) | 35.0 (21.6) | [0.000] | [0.001] |
| Low-skilled Offices | 28.4 (19.7) | 23.5 (18.9) | 18.9 (16.8) | [0.000] | [0.000] |
| Local Services | 18.3 (15.3) | 15.6 (14.6) | 11.6 (12.2) | [0.000] | [0.000] |
| Non-offices | 32.6 (21.6) | 29.3 (19.7) | 26.9 (18.9) | [0.001] | [0.036] |
| Observations | 67 | 135 | 276 | | |

*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.

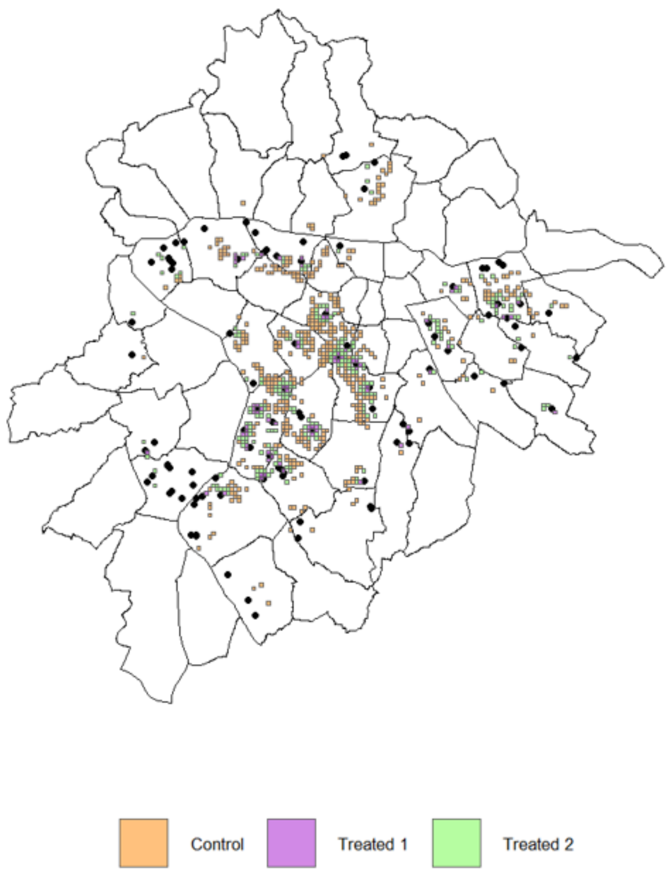
Table D4. Effects of New Commercial Buildings: No Matching

| | Log Estabs (1) | Log Workers (2) | % College (3) | Wage Premium (4) |
|---|---|---|---|---|
| **Panel A. High-Skilled Offices** | | | | |
| 0-250m | 0.0665* | 0.1849** | 0.0233* | 0.0295 |
| | (0.0396) | (0.0820) | (0.0140) | (0.0260) |
| $R^2$ | 0.00521 | 0.00917 | 0.00452 | 0.00290 |
| Observations | 5,145 | 5,145 | 5,145 | 5,145 |
| 250-500m | 0.0377 | 0.0824 | -0.0022 | -0.0010 |
| | (0.0277) | (0.0640) | (0.0091) | (0.0149) |
| $R^2$ | 0.00201 | 0.00205 | -0.00011 | -0.00016 |
| Observations | 6,165 | 6,165 | 6,165 | 6,165 |
| **Panel B. Low-Skilled Offices** | | | | |
| 0-250m | 0.0668* | 0.0491 | 0.0014 | -0.0437** |
| | (0.0373) | (0.1103) | (0.0125) | (0.0197) |
| $R^2$ | 0.00633 | 0.00025 | -0.00017 | 0.00818 |
| Observations | 5,145 | 5,145 | 5,145 | 5,145 |
| 250-500m | 0.0418 | 0.0668 | -0.0083 | -0.0061 |
| | (0.0265) | (0.0714) | (0.0094) | (0.0136) |
| $R^2$ | 0.00315 | 0.00096 | 0.00082 | 0.00006 |
| Observations | 6,165 | 6,165 | 6,165 | 6,165 |
| **Panel C. Local Services** | | | | |
| 0-250m | 0.0649* | 0.1038* | -0.0111* | -0.0197* |
| | (0.0360) | (0.0545) | (0.0063) | (0.0110) |
| $R^2$ | 0.00938 | 0.00870 | 0.00432 | 0.00464 |
| Observations | 5,145 | 5,145 | 5,145 | 5,145 |
| 250-500m | -0.0103 | -0.0188 | -0.0117** | -0.0202** |
| | (0.0170) | (0.0307) | (0.0053) | (0.0084) |
| $R^2$ | 0.00028 | 0.00028 | 0.00567 | 0.00598 |
| Observations | 6,165 | 6,165 | 6,165 | 6,165 |
| **Panel D. Non-Offices** | | | | |
| 0-250m | 0.0108 | 0.0675 | 0.0167 | 0.0057 |
| | (0.0426) | (0.0787) | (0.0109) | (0.0190) |
| $R^2$ | 0.00003 | 0.00137 | 0.00308 | 0.00005 |
| Observations | 5,145 | 5,145 | 5,145 | 5,145 |
| 250-500m | -0.0095 | -0.0432 | -0.0086 | -0.0107 |
| | (0.0232) | (0.0475) | (0.0083) | (0.0126) |
| $R^2$ | 0.00003 | 0.00074 | 0.00094 | 0.00053 |
| Observations | 6,165 | 6,165 | 6,165 | 6,165 |

*Notes:* This table reports estimates of new building effects for different outcome variables without employing matching, as described in Section 6.3. Standard errors clustered at the cell level are displayed in parentheses. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

**Effects of New Residential Buildings**

Figure D2. Empirical Analysis Setup




*Notes:* This figure depicts the design of the empirical analysis using residential buildings as the shock. Details are provided in Section 6.4.

Table D5. Summary Statistics: New Residential Buildings

| | Median | Mean | Std. Deviation | Min | Max |
|---|---|---|---|---|---|
| Total Land Area ($m^2$) | 9,561.5 | 11,144.2 | 5,246.8 | 5,363 | 26,620 |
| Occupied Land Area ($m^2$) | 6,079 | 6,726.1 | 3,229.2 | 2,121 | 15,959 |
| Built-Area-Ratio | 9.4 | 11.2 | 5.6 | 4.6 | 30.5 |

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

Table D6. Effects of New Residential Buildings

| | Log Estabs<br>(1) | Log Workers<br>(2) | % College<br>(3) | Wage Premium<br>(4) |
|---|---|---|---|---|
| **Panel A. High-Skilled Offices** | | | | |
| 0-250m | 0.0155 | 0.0977 | -0.0280* | 0.0095 |
| | (0.0394) | (0.0896) | (0.0145) | (0.0223) |
| $R^2$ | 0.00008 | 0.00165 | 0.00440 | 0.00012 |
| Observations | 7,770 | 7,770 | 7,770 | 7,770 |
| 250-500m | -0.0190 | -0.0027 | 0.0014 | -0.0247* |
| | (0.0232) | (0.0489) | (0.0078) | (0.0137) |
| $R^2$ | 0.00042 | -0.0001 | -0.00008 | 0.00270 |
| Observations | 10,005 | 10,005 | 10,005 | 10,005 |
| **Panel B. Low-Skilled Offices** | | | | |
| 0-250m | 0.0441 | -0.0474 | -0.0119 | -0.0323 |
| | (0.0346) | (0.0922) | (0.0124) | (0.0202) |
| $R^2$ | 0.00218 | 0.00019 | 0.00113 | 0.00359 |
| Observations | 7,770 | 7,770 | 7,770 | 7,770 |
| 250-500m | 0.0232 | 0.0128 | -0.0065 | 0.0059 |
| | (0.0227) | (0.0613) | (0.0078) | (0.0115) |
| $R^2$ | 0.00089 | -0.00006 | 0.00052 | 0.00011 |
| Observations | 10,005 | 10,005 | 10,005 | 10,005 |
| **Panel C. Local Services** | | | | |
| 0-250m | 0.0563 | 0.0924 | 0.0043 | 0.0147 |
| | (0.0400) | (0.0665) | (0.0074) | (0.0134) |
| $R^2$ | 0.00568 | 0.00499 | 0.00033 | 0.00168 |
| Observations | 7,770 | 7,770 | 7,770 | 7,770 |
| 250-500m | -0.0093 | 0.0060 | -0.0077* | -0.0033 |
| | (0.0145) | (0.0261) | (0.0046) | (0.0081) |
| $R^2$ | 0.00026 | -0.00005 | 0.00237 | 0.00005 |
| Observations | 10,005 | 10,005 | 10,005 | 10,005 |
| **Panel D. Non-Offices** | | | | |
| 0-250m | 0.0365 | -0.0008 | 0.0174* | 0.0201 |
| | (0.0400) | (0.0712) | (0.0101) | (0.0169) |
| $R^2$ | 0.00146 | -0.00013 | 0.00290 | 0.00150 |
| Observations | 7,770 | 7,770 | 7,770 | 7,770 |
| 250-500m | 0.0082 | -0.0198 | -0.0037 | -0.0091 |
| | (0.0185) | (0.0390) | (0.0066) | (0.0118) |
| $R^2$ | 0.00005 | 0.00010 | 0.00014 | 0.00041 |
| Observations | 10,005 | 10,005 | 10,005 | 10,005 |

*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):

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

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

Table D7. Effects of New Commercial Buildings: Continuous Treatment

| | Log Estabs (1) | Log Workers (2) | % College (3) | Wage Premium (4) |
|---|---|---|---|---|
| **Panel A. High-Skilled Offices** | | | | |
| $(1 - dist)T$ | 0.0899 | 0.2362* | 0.0410** | 0.0816** |
| | (0.0607) | (0.1261) | (0.0194) | (0.0325) |
| $R^2$ | 0.90336 | 0.83397 | 0.62490 | 0.68715 |
| Observations | 5,130 | 5,130 | 5,130 | 5,130 |
| **Panel B. Low-Skilled Offices** | | | | |
| $(1 - dist)T$ | 0.0638 | 0.1187 | -0.0271 | -0.0161 |
| | (0.0549) | (0.1566) | (0.0192) | (0.0305) |
| $R^2$ | 0.91065 | 0.77187 | 0.53717 | 0.64140 |
| Observations | 5,130 | 5,130 | 5,130 | 5,130 |
| **Panel C. Local Services** | | | | |
| $(1 - dist)T$ | 0.1024** | 0.1027 | -0.0225** | -0.0110 |
| | (0.0488) | (0.0769) | (0.0101) | (0.0180) |
| $R^2$ | 0.93529 | 0.89294 | 0.75306 | 0.75935 |
| Observations | 5,130 | 5,130 | 5,130 | 5,130 |
| **Panel D. Non-Offices** | | | | |
| $(1 - dist)T$ | 0.0069 | 0.0161 | 0.0127 | -0.0386 |
| | (0.0532) | (0.1011) | (0.0170) | (0.0270) |
| $R^2$ | 0.86761 | 0.80721 | 0.70358 | 0.73987 |
| Observations | 5,130 | 5,130 | 5,130 | 5,130 |

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

## SE Clustering using Closest New Building

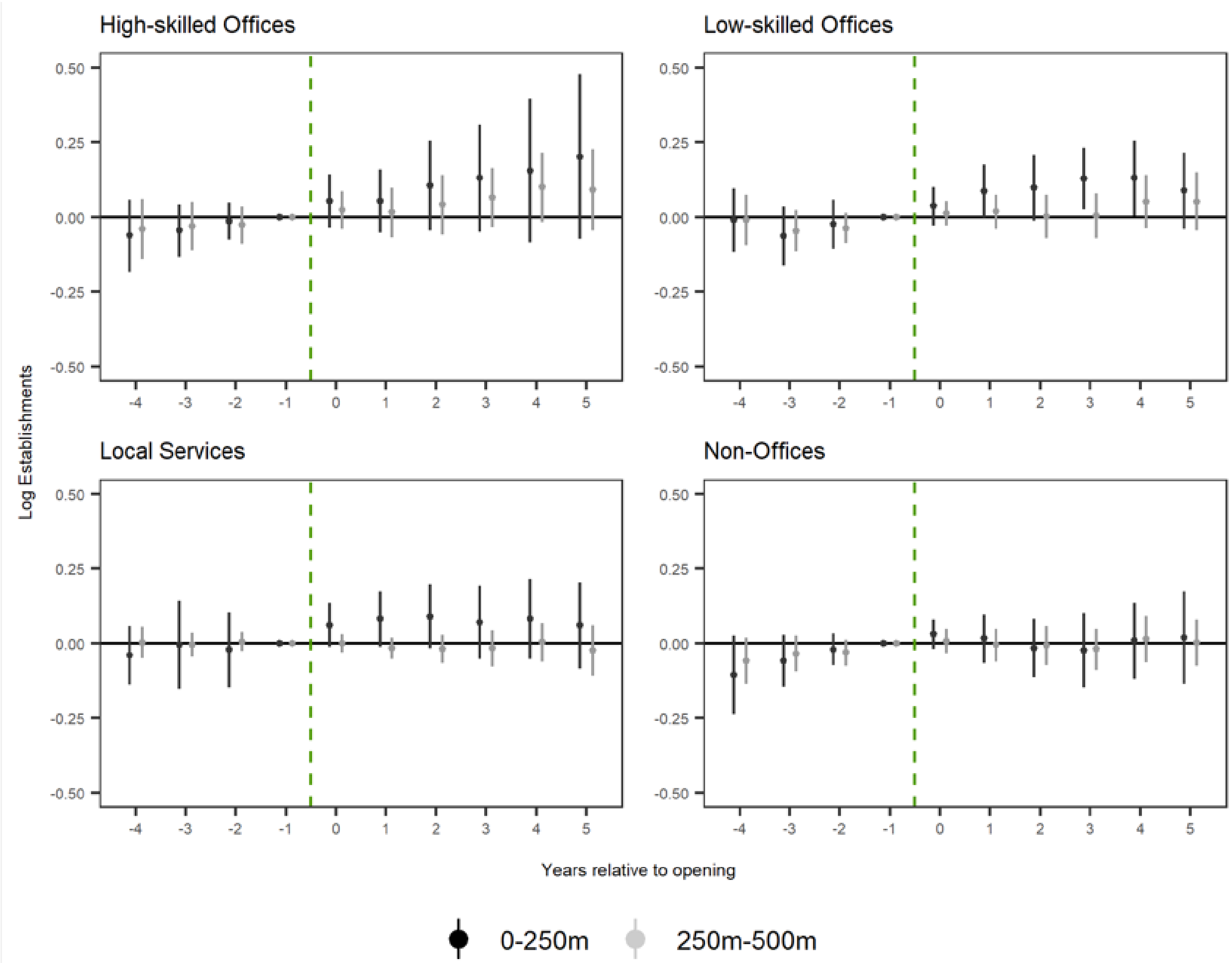
Table D8. Effects of New Commercial Buildings: Alternative Clustering

| | Log Estabs | Log Workers | % College | | Wage Premium | |
|---|---|---|---|---|---|---|
| | | | All estabs | excl. new estabs | All estabs | excl. new estabs |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| **Panel A. High-Skilled Offices** | | | | | | |
| 0-250m | 0.0860* | 0.1994** | 0.0296** | 0.0216* | 0.0539** | 0.0150 |
| | (0.0470) | (0.0830) | (0.0142) | (0.0129) | (0.0224) | (0.0201) |
| $R^2$ | 0.01148 | 0.01668 | 0.01293 | 0.00799 | 0.01351 | 0.00110 |
| Observations | 1,920 | 1,920 | 1,920 | 1,830 | 1,920 | 1,830 |
| 250-500m | 0.0168 | 0.0552 | 0.0014 | -0.0073 | 0.0129 | -0.0054 |
| | (0.0324) | (0.0812) | (0.0094) | (0.0107) | (0.0167) | (0.0171) |
| $R^2$ | 0.00022 | 0.00078 | -0.00022 | 0.00041 | 0.00061 | -0.00011 |
| Observations | 4,020 | 4,020 | 4,020 | 3,645 | 4,020 | 3,645 |
| **Panel B. Low-Skilled Offices** | | | | | | |
| 0-250m | 0.0747 | 0.1481 | -0.0016 | -0.0214* | -0.0256 | -0.0380 |
| | (0.0527) | (0.1166) | (0.0144) | (0.0120) | (0.0280) | (0.0284) |
| $R^2$ | 0.01040 | 0.00483 | -0.00049 | 0.00607 | 0.00326 | 0.00878 |
| Observations | 1,920 | 1,920 | 1,920 | 1,770 | 1,920 | 1,770 |
| 250-500m | 0.0414 | 0.1038 | -0.0126 | -0.0218** | -0.0090 | -0.0545*** |
| | (0.0321) | (0.0713) | (0.0125) | (0.0106) | (0.0152) | (0.0135) |
| $R^2$ | 0.00333 | 0.00266 | 0.00191 | 0.00657 | 0.00026 | 0.01952 |
| Observations | 4,020 | 4,020 | 4,020 | 3,780 | 4,020 | 3,780 |
| **Panel C. Local Services** | | | | | | |
| 0-250m | 0.0833** | 0.1434** | -0.000005 | -0.0107 | -0.0074 | -0.0325** |
| | (0.0417) | (0.0693) | (0.0077) | (0.0090) | (0.0149) | (0.0156) |
| $R^2$ | 0.01516 | 0.01982 | -0.00052 | 0.00353 | 0.00040 | 0.01181 |
| Observations | 1,920 | 1,920 | 1,920 | 1,905 | 1,920 | 1,905 |
| 250-500m | 0.0139 | -0.0019 | -0.0097 | -0.0117* | -0.0163* | -0.0223** |
| | (0.0223) | (0.0427) | (0.0067) | (0.0062) | (0.0097) | (0.0106) |
| $R^2$ | 0.00060 | -0.00024 | 0.00381 | 0.00628 | 0.00382 | 0.00737 |
| Observations | 4,020 | 4,020 | 4,020 | 3,945 | 4,020 | 3,945 |
| **Panel D. Non-Offices** | | | | | | |
| 0-250m | -0.0170 | 0.0967 | 0.0270* | 0.0260* | 0.0102 | 0.0004 |
| | (0.0668) | (0.0958) | (0.0141) | (0.0138) | (0.0225) | (0.0180) |
| $R^2$ | 0.00004 | 0.00451 | 0.01278 | 0.01088 | 0.00017 | -0.00054 |
| Observations | 1,920 | 1,920 | 1,920 | 1,845 | 1,920 | 1,845 |
| 250-500m | -0.0341 | -0.0715 | 0.0049 | 0.0130 | -0.0041 | -0.0311** |
| | (0.0269) | (0.0564) | (0.0098) | (0.0093) | (0.0127) | (0.0123) |
| $R^2$ | 0.00248 | 0.00251 | 0.00013 | 0.00211 | -0.00014 | 0.00534 |
| Observations | 4,020 | 4,020 | 4,020 | 3,870 | 4,020 | 3,870 |

*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.

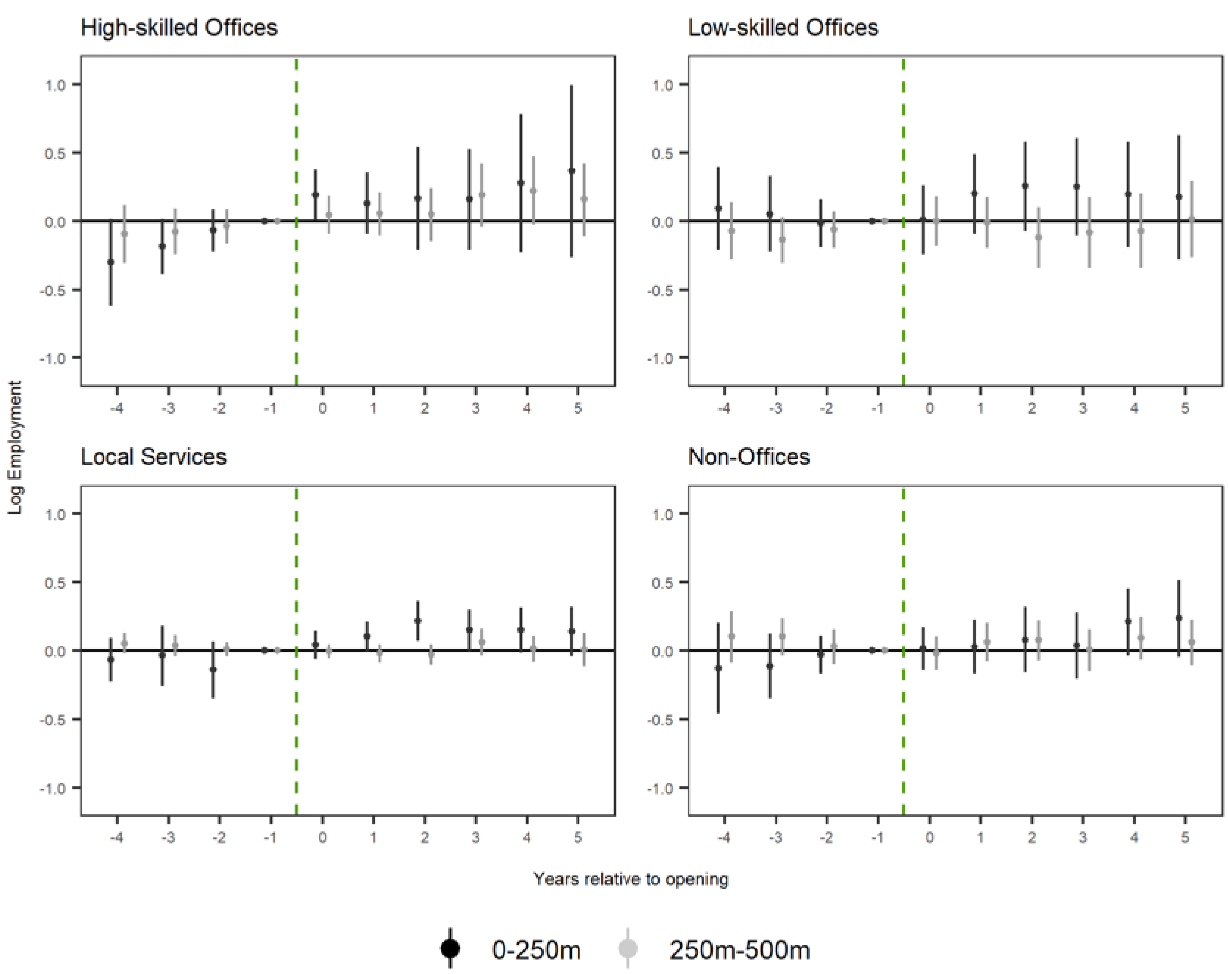
### Callaway and Sant'Anna (2021)'s Estimator

Figure D3. Event Study using Callaway and Sant'Anna (2021): Number of Establishments by Sector



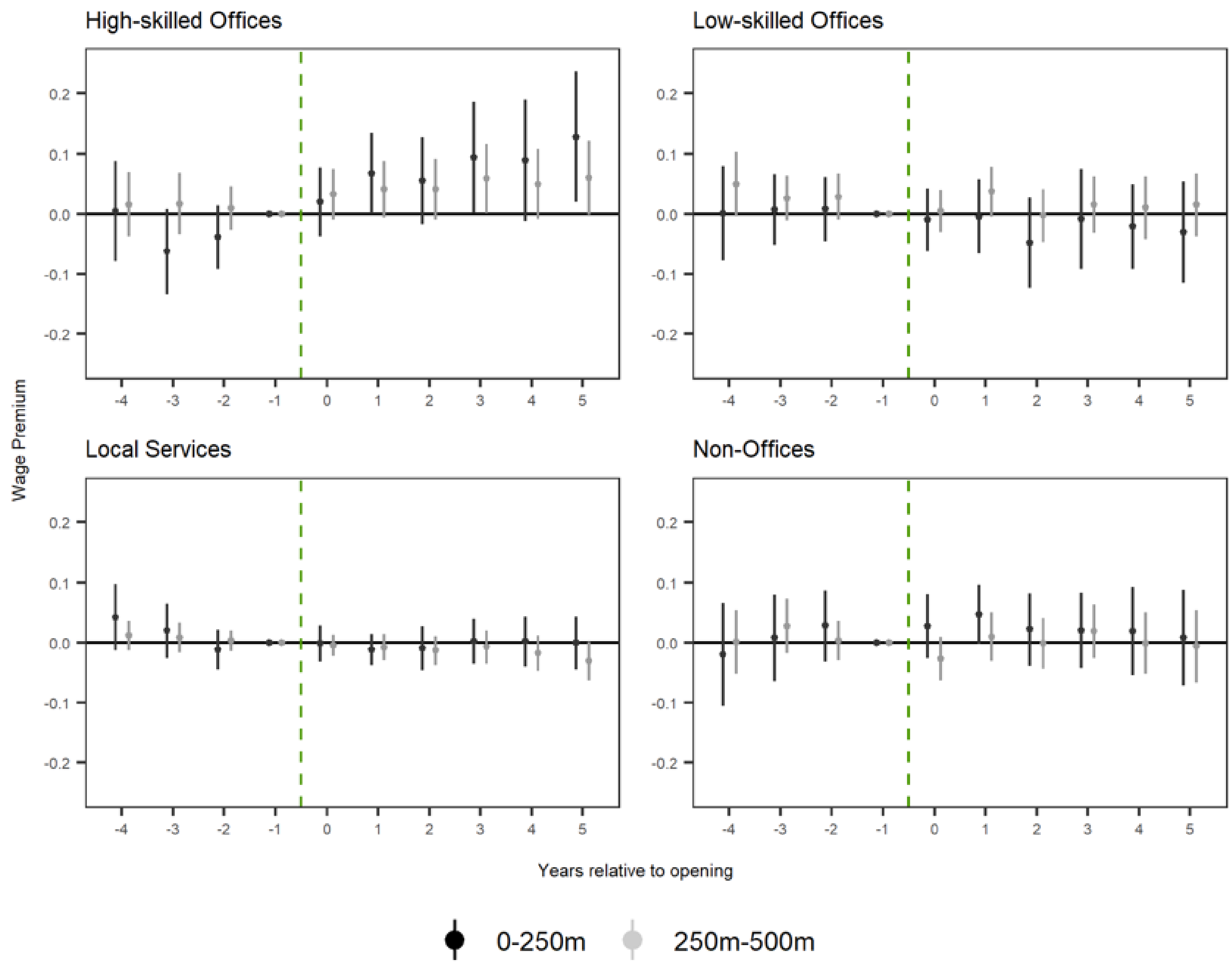

*Note:* This figure plots coefficients from Callaway and Sant'Anna (2021)'s estimator using the log number of establishments for different sectors as the outcome variable. The definition of each sector is described in Section 2.2. The bars indicates the 95% confidence interval, where standard errors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

Figure D4. Event Study using Callaway and Sant'Anna (2021): Employment by Sector



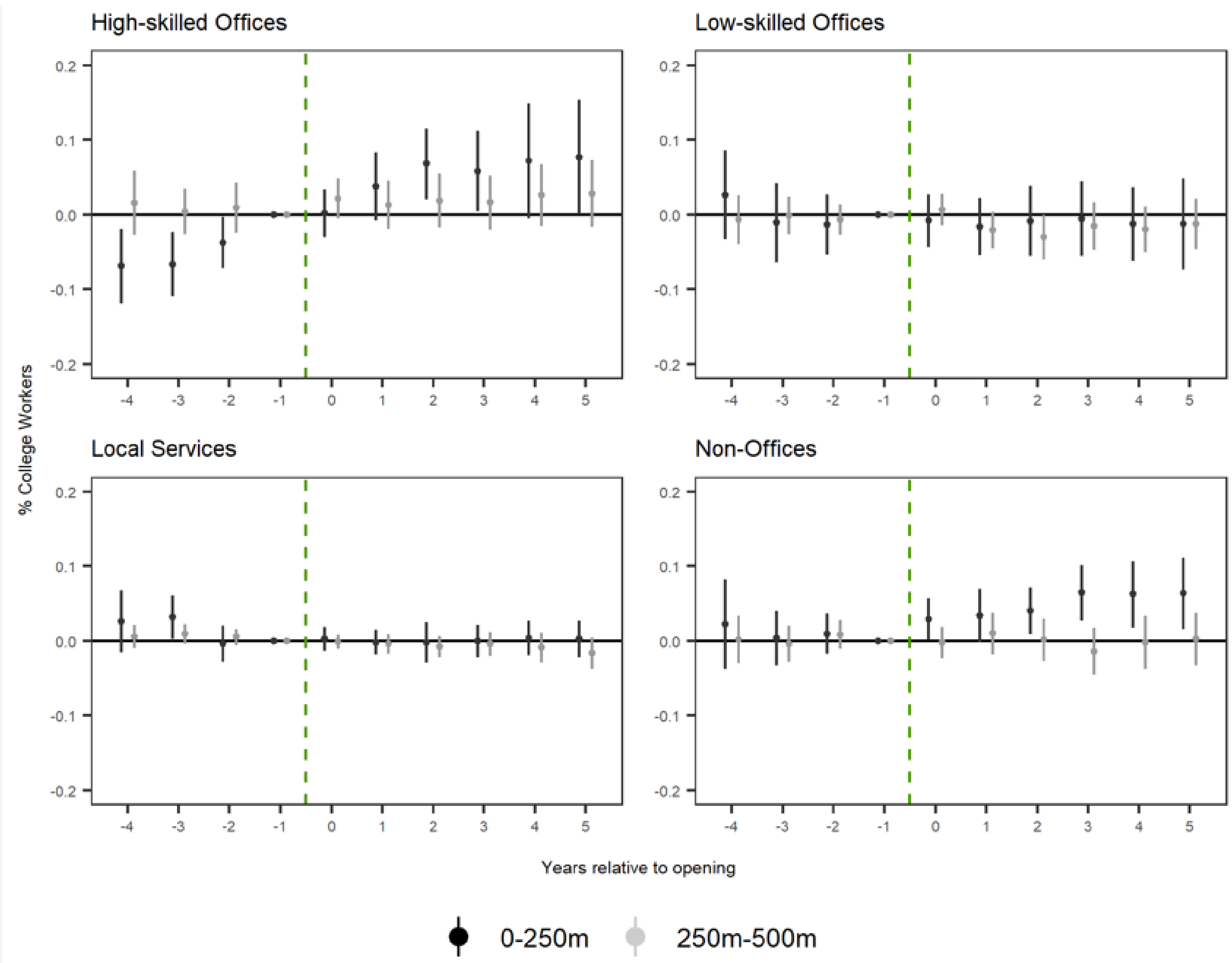

*Note:* This figure plots coefficients from Callaway and Sant'Anna (2021)'s estimator using log employment for different sectors as the outcome variable. The definition of each sector is described in Section 2.2. The bars indicates the 95% confidence interval, where standard errors are clustered at the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

Figure D5. 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 D6. 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.