The spatial scope of agglomeration economies in Brazil*

Edilberto T. Almeida† Raul M. Silveira Neto† Roberta M. Rocha†

Abstract

This paper provides evidences about the location and colocation patterns of the new firms of manufacturing and the spatial scope of agglomeration economies in Brazil. Using unique geocoded data for all Brazilian manufacturing activities and distance-based measures, we find that there are patterns of colocalization between entrants and existing establishments and that these patterns occurs mainly at short distances. Furthermore, exploring the spatial scope of the agglomeration economies, focusing on the localization effects (own-industry employment), we find that these effects, particularly to industries colocalized at short distances, attenuate rapidly with distance (around 5 km). These results are robust to the inclusion of a comprehensive set of controls variables for observable and unobservable local characteristics and the use of instrumental variables to address endogeneity concerns.

Keywords: attenuation, geocoded data, distance-based measures, Brazilian manufacturing industries

Resumo

Este artigo fornece evidências sobre os padrões de localização e colocalização das novas firmas de manufatura e o escopo espacial das economias de aglomeração no Brasil. Usando dados georreferenciados para a indústria da transformação no Brasil e medidas baseadas em distâncias, os resultados mostram que há um padrão de colocalização entre os novos estabelecimentos e os estabelecimentos existentes, principalmente a curtas distâncias. Os resultados também mostram que os efeitos das economias de aglomeração gerados pela concentração do emprego na própria indústria, principalmente nos setores colocalizados a curtas distâncias, são rapidamente atenuados com a distância (em torno de 5 km). Estes resultados são robustos a inclusão de um amplo conjunto de variáveis de controle para características locais observáveis e não observáveis e ao uso de variáveis instrumentais.

Palavras-chave: atenuação, dados georreferenciados, medidas baseadas em distâncias, indústria da transformação no Brasil

JEL classification: L26, L60, R12

*The authors acknowledge and are grateful for financial support by the Conselho Nacional de Desenvolvimento Científico e Tecnológico CNPq.

†Federal University of Pernambuco - UFPE. E-mail: edilbertotiago@hotmail.com.

1 Introduction

Agglomeration economies are one of the reasons why cities offer better jobs and provide attractive environments for more productive firms and new establishments (Marshall, 1890; Carlton, 1983; Head et al., 1995). Together with other local factors,[1] a better understanding of the spatial scope of these externalities sheds light on why some places are more entrepreneurial than others (Glaeser and Kerr, 2009; Duranton, 2015). Given these issues and the political relevance of the theme to regional and urban development, the relationship between new firm location choices and the agglomeration economies has been widely studied (see, e.g., Arauzo-Carod et al., 2010, for a survey). Most of these studies, however, use aggregated geographic data and do not capture microgeographic patterns that occur within cities, for example at the neighborhood level, because they assume implicitly that the agglomeration economies operate homogeneously within cities.

This paper addresses an important question still little studied, especially in developing countries, namely the spatial scope of agglomeration economies. As raised by Rosenthal and Strange (2003), what is the geographic and industrial scope of agglomeration externalities? Empirical evidence for developed countries shows that the agglomeration economies tend to be attenuated with distance. Using creation of and employment by new establishments, Rosenthal and Strange (2003) found that agglomeration economies, particularly the localization effects (own-industry employment), are attenuated around 10 km in the US. Based on other outcome variables, such as wages (see, e.g., Fu, 2007; Rosenthal and Strange, 2008; and more recently Håkansson and Isacsson, 2019 for the US and Sweden) or TFP (such as Andersson et al. (2019) for the Sweden), the results are generally similar. In the context of developing economies, only Li et al. (2020) provide evidence about the spatial scope of agglomeration economies (for China). The authors found that the effects of localization economies are attenuated more rapidly than in developed countries. These results suggest that the spatial scope of agglomeration economies is different in developing countries, which can be related to the quality of urban infrastructure.

Besides the scarcity of detailed evidence on the subject, some economic characteristics also make study of the spatial scope of agglomeration economies particularly interesting in Brazil. For example, unlike China, historically there has been no restriction on worker mobility, and economic activities are more market oriented in Brazil, which can substantially affect the geographical distribution of activities. But little is known about this phenomenon. Previous works are based exclusively on between-city variation in the data (see, e.g., Barufi et al., 2016; Chauvin et al., 2017). Still from this perspective, recent evidence shows that manufacturing activity in Brazil is more concentrated than in other developing countries, such as China and Russia, and much more concentrated than in developed countries. In addition, a better understanding of the spatial scope of these externalities is of unquestionable political relevance. Local development public policies to attract new establishments have been on the agenda of local governments for decades in developing countries, particularly in Brazil (Leff, 1972; Hansen, 1987; Tatsch et al., 2015). Most of these policies, however, are not based on detailed studies of intrinsic market factors such as economies of agglomeration. Instead, they are only supported by a broad range of political interests, and are not always economically efficient (Varsano, 1997; Paes and Siqueira, 2005; 2008).

Here we seek to fill part of this gap in the literature. For this, we use a geocoded employer-employee database of Brazilian manufacturing activities. Initially, to better understand the pattern of location of manufacturing entrepreneurship, we use the nonparametric approach developed by Duranton and Overman (2005; 2008) to document both location and colocation patterns of new manufacturing plants, considering in particular the location of new entrants versus existing establishments. This preliminary data investigation provides insight about geographic proximity between entrants and incumbents establishments and identifies the industrial sectors for which these patterns are most prominent. In turn, these patterns may be associated with the entrants' locational choice due to local externalities generated by proximity to existing establishments.

In this context, taking advantage of characteristics of our database, we use exogenously defined microgeographic areas instead of the official administrative areas to examine the spatial extent of agglomeration

[1] Such as local industry structure, demographics, scale economies, and cost advantages associated with city characteristics (see, e.g., Glaeser, 2007; Glaeser and Kerr, 2009; Glaeser et al., 2010a).

economies on the location decisions of new establishments and on the employment levels that they choose. Specifically, we estimate the local determinants of the number of firm births per square kilometer and their associated employment levels as functions of the own-industry employment and other economic environment characteristics. Although our focus is on new establishments, that choose the locations by taking the existing economic environment as given, unobserved characteristics that affect both existing business concentrations and attract new establishments make our estimates inconsistent.

To address these potential concerns, we use different tools that involve both the wealth of detail in our data and techniques to deal with the presence of endogenous explanatory variables in nonlinear models. Different from previous studies, such as Rosenthal and Strange (2003) and Li et al. (2020), we have panel data which allows us to control for any observed and unobserved heterogeneities fixed in time in different neighborhoods (or districts) within cities, and therefore allow comparing areas of one square kilometer within the same neighborhood. In this sense, our microgeographic areas of one square kilometer are generally outside of the firm's set of choices because they depend on land availability and use, minimizing potential selection problems. We also include a comprehensive set of control variables for economic environment, previously existing transportation infrastructure, geographic characteristics, and local development policies about the sites chosen by the new establishments. In addition, to address any remaining source of heterogeneity, we use a control function approach with a shift-share instrumental variable that exploits the changes in national employment growth specific to the industry to generate exogenous variation at the microgeographic area level.

The main result here shows that agglomeration economies are attenuated with distance. In particular, the effect of own-industry employment at 1 km is significantly larger than the effect of employment further away, indicating that initial attenuation is rapid. For example, adding 100 workers in the same industry up to 1 km would generate, on average, an increase of 16.8% in the expected number of births and 30% in the expected number of employees. From this same perspective, adding 100 additional employees to the 1-5 km ring would result, on average, an increase of 2.6% in the expected number of births and 10% in the expected number of employees. On the other hand, in nearly all cases for both births and new establishment employment, localization effects disappear after 5 km. The pattern of attenuation with distance remains largely robust to the inclusion of different control variables and the use of instrumental variables, strengthening the reliability of our estimates. Our results for Brazil are consistent with theoretical models of urban areas and previous empirical evidences evidence from other countries.

The paper is organized as follows. Section 2 presents our data source and briefly discusses the location patterns of new establishments obtained from Duranton and Overman (2005; 2008)'s measures. Section 3 presents the empirical approach based on count models of births and new-establishment employment. Section 4 discusses and compares the results them with the evidence obtained in other studies. Section 5 concludes.

2 Data and spatial location of new establishments

2.1 Data

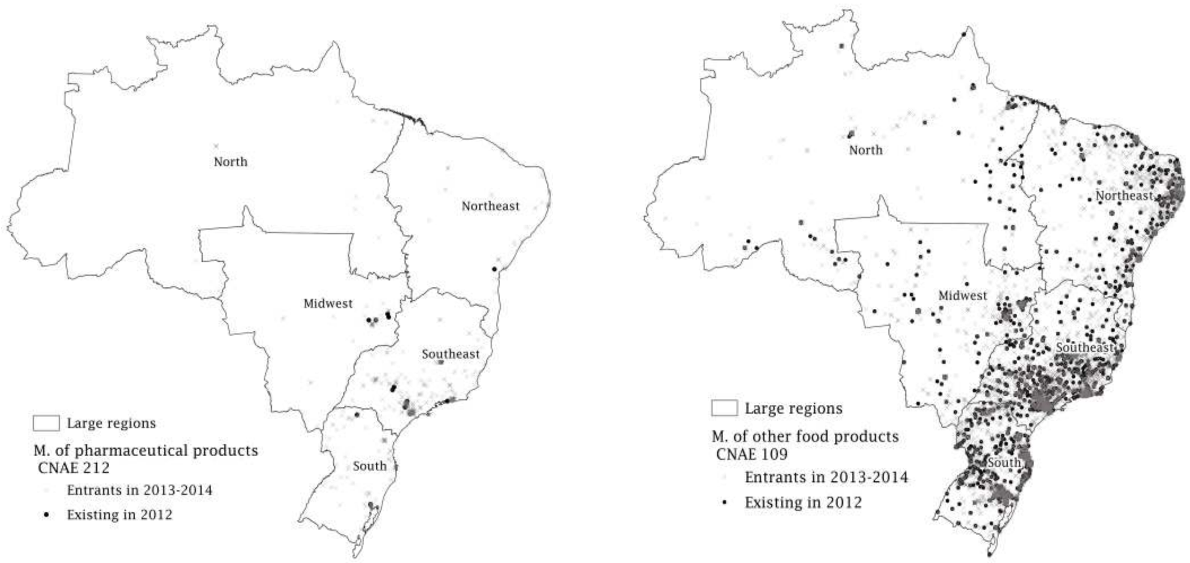
We use an exhaustive establishment-level dataset from the Annual Report of Social Information (*Relação Anual de Informações Sociais*, or RAIS), made available annually by the Ministry of Labor and Employment. This database encompasses all formal establishments in Brazil. In this database, each establishment has a unique identifier, the number on the National Registry of Legal Entities (CNPJ). The data include firm address, date of opening (and closing, if applicable), number of active jobs, and the National Classification of Economic Activities (CNAE) version 2.0 (which is compatible with the International Standard Industrial Classification of all Economic Activities (ISIC) revision 4). Using these data for the period 2007-2014, geocoded every year[2], allows us to assess in detail where the new manufacturing establishments in Brazil were spatially located during the period studied. Particularly, we can divide establishments into

[2]For each year, more than 99% of new establishments were geocoded. Establishments can change address over the years, so we consider each establishment's birth address to obtain the geographical coordinates.

new entrants and existing ones in each year.

Reflecting the general location pattern of the manufacturing industry (see, e.g., Silveira Neto, 2005; Lautert and Araújo, 2007; Rocha et al., 2019), the new enterprises are concentrated mainly in the Southeast and South regions. Both in 2007-2008 (75.18%) and 2013-2014 (71.66%) the two regions concentrated more than 70% of the new establishments.[3] This was expected, given that entrepreneurial activity occurs more frequently in the most dynamic regions, where more business opportunities are present. At a smaller geographic scale, the state of São Paulo represents 23.44% and 20.87% of the total of entrants in the country in the two periods, while the São Paulo Metropolitan Region (SPMR) represents 9.83% and 7.68% of the total of entrants in the country in the same period. Note also that investments are, on average, larger in São Paulo (measured by the number of workers) as a result of the larger market potential.

Based on data broken down by industry, the location patterns of new establishments are also heterogeneous. In this context, another interesting way to visualize this heterogeneity is to look at the location of entrants relative to existing establishments. Figures 1 (a) and (b) show these locations for two illustrative industries (entrants in 2013-2014 relative to existing establishments in 2012). Figure 1 (a) depicts the entrants (cross) and existing establishments (circle) engaged in the *manufacture of pharmaceutical products* - CNAE 212. Similarly, in Figure 1 (b) shows the entrants and incumbents involved in the *manufacture of other food products* - CNAE 109. A careful look suggests that the entrants in CNAE 212 appear to be more concentrated than existing establishments. On the other hand, for entrants in CNAE 109, there does not appear to be a greatly different pattern of spatial distribution between the entrants and incumbents.



(a) M. of pharmaceutical products - CNAE 212

(b) M. of other food products - CNAE 109

Figure 1. Maps of two illustrative industries

While these preliminary remarks are useful for examining visibly clear patterns, more specific relationships, such as colocalization between entrants and existing establishments, are not easily identified. In the next subsection we rigorously explore location and colocation patterns using distance-based measures.

[3] We calculate the number of new establishments every two years. There are technical and economic criteria that justify this division. Technically, as will become clear in the next subsection, some industries have few new establishments annually, which poses a limitation to the use of distance-based measures (Duranton and Overman, 2005; 2008; Klier and McMillen, 2008). Economically, some forces operating in the economic environment around the new plants (e.g., Marshallian agglomeration forces) do not vary significantly from one year to the next.

2.2 Location patterns of new establishments

To better understand the geographical distribution of manufacturing entrepreneurship,[4] we use the measures developed by Duranton and Overman (2005; 2008) for localization and colocalization, weighted by employment, to assess the location patterns of new establishments for each industry at the 3-digit level. This is a distance-based method, so it is not susceptible to the modifiable areal unit problem (MAUP) common in other traditional measures of concentration (e.g., Gini index and Ellison and Glaeser (1997)'s index). Specifically, from the continuous localization measure, we can assess, for example, whether new establishments follow the same location pattern, and whether they are localized or dispersed compared to existing establishments in the same industry.[5] On the other hand, from the colocalization measure, we can assess whether new establishments are colocalized relative to existing establishments, i.e., if entrants locate near to (or far from) incumbents.

Both localization and colocalization measures are obtained from bilateral Euclidean distances. In the first, we consider the bilateral distances between all entrants of a specific industry, while in the second, we consider the bilateral distances between each entrant and all existing establishments in the previous period. To illustrate the logic of this measure, we provide some illustrative examples. We begin by presenting examples from localization measurements. The black solid lines in Figures 2 (a) and (c) plot the K-density estimates for the entrants in 2013-2014 relative to existing establishments in 2012 in the industries of *manufacture of pharmaceutical products* - CNAE 212 and *manufacture of other food products* - CNAE 109. Graphically, one can detect whether new establishments are localized relative to existing ones when the K-density lies above its upper confidence band (delimited by the extremes of the hatched area that determines the confidence interval containing 95% of counterfactual distributions). On the other hand, we consider that the entrants are dispersed relative to the incumbents when the K-density lies below its lower confidence band for some distance and never exceeds the upper confidence band. When the K-density is within the confidence interval, we can assume that the entrants do not follow a pattern of spatial distribution different from the existing establishments. As can be seen in Figures 2 (a) and (c), the impression from observing the maps in Figures 1 (a) and (b) is confirmed. Note, for example, that the entrants in the *manufacture of pharmaceutical products* are localized relative to the industry, while the entrants in the *manufacture of other food products* have location patterns similar to existing establishments.

As discussed earlier, another interesting issue is to consider the colocation patterns between entrants and existing establishments. As in Duranton and Overman (2008), we also provide evidence about this. For this study, this pattern is particularly important, since our focus is on the spatial scope of agglomeration economies and therefore is directly related to the proximity between plants in the same industry. Figures 2 (b) and (d) plot the K-density estimates of bilateral distances between entrants and all existing establishments for the same industries. Colocation and codispersion can be detected by proceeding as before, i.e., looking at K-density estimates in the black solid lines. Two interesting patterns emerge. First, as examples of colocalization, the entrants in *manufacture of pharmaceutical products* and *manufacture of other food products* are colocalized with existing establishments. Second, the colocalization occurs at short distances.

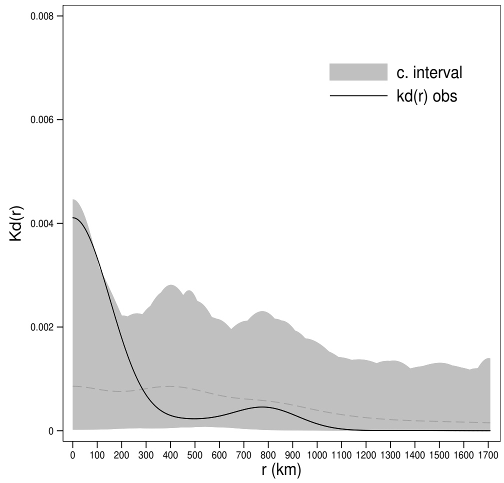
We can also get an overview for all manufacturing activity. We start with 103 industries at the 3-digit level in each period (2007-2008 and 2013-2014), and as in Duranton and Overman (2008), we drop 16 and 25 industries with fewer than 10 entrants in the first and second periods, respectively. Among the remaining establishments, 14.94% and 12.82% of employment-weighted entrants are localized in each period,[6] as can be seen in Table 1. In contrast, 8.05% and 10.26% of entrants are dispersed, while for most industries (around 77% in both periods), entrants do not have statistically different location patterns from those observed for existing establishments. Interestingly, when we look at these percentages by distance, there is

[4]There is evidence that cities with younger plants and more entrepreneurship have higher growth rates, thus suggesting that local entrepreneurship is important for economic development (see, e.g., Faberman, 2011; Glaeser et al., 2015; Fritsch and Wyrwich, 2016). In this context, it is worthwhile investigating in more detail the geographic location patterns of new manufacturing plants.

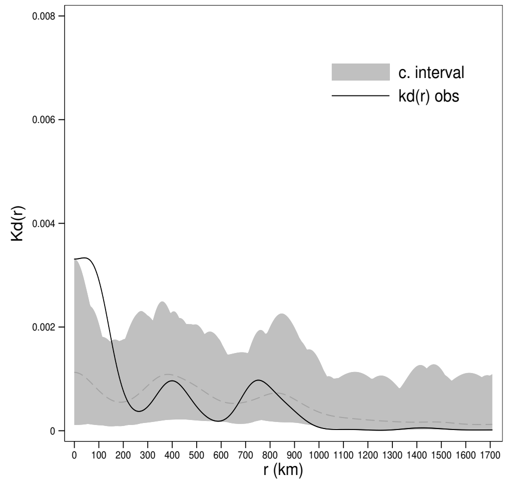
[5]As in Duranton and Overman (2008), we define as counterfactual the locations that contain establishments of the same industry only. Due to space limitations, we have not included the Appendices with detailed information about methodology. This information can be made available upon request.

[6]These percentages are similar those found by Duranton and Overman (2008) for the UK (13%).

no specific pattern of localization at short distances, although there are differences in the patterns between the two periods (see Figure 3 (a)). In general, this evidence for Brazil is similar to that found for the UK and indicates that there is no clear tendency for manufacturing activities to become systematically more or less clustered over time because of new establishments.

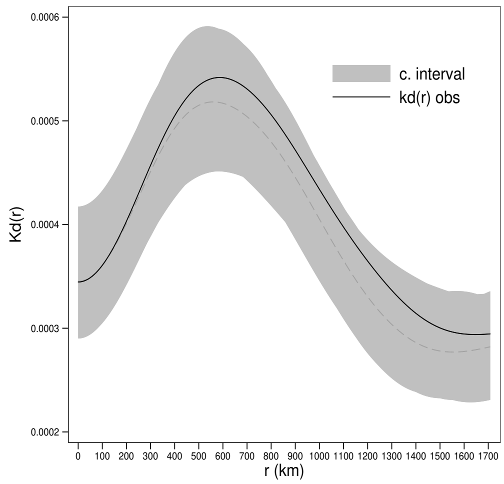


(a) Localization

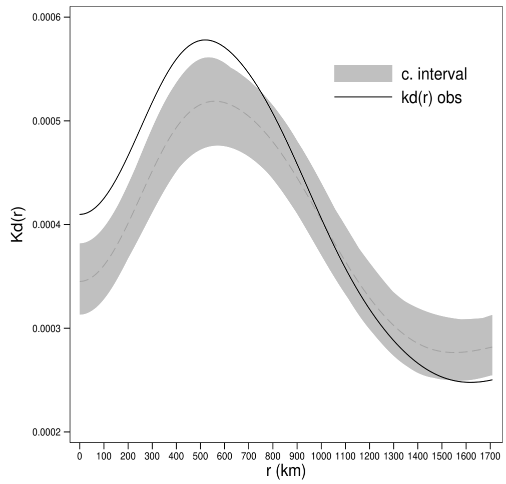


(b) Colocalization

M. of pharmaceutical products - CNAE 212



(c) Localization



(d) Colocalization

M. of other food products - CNAE 109

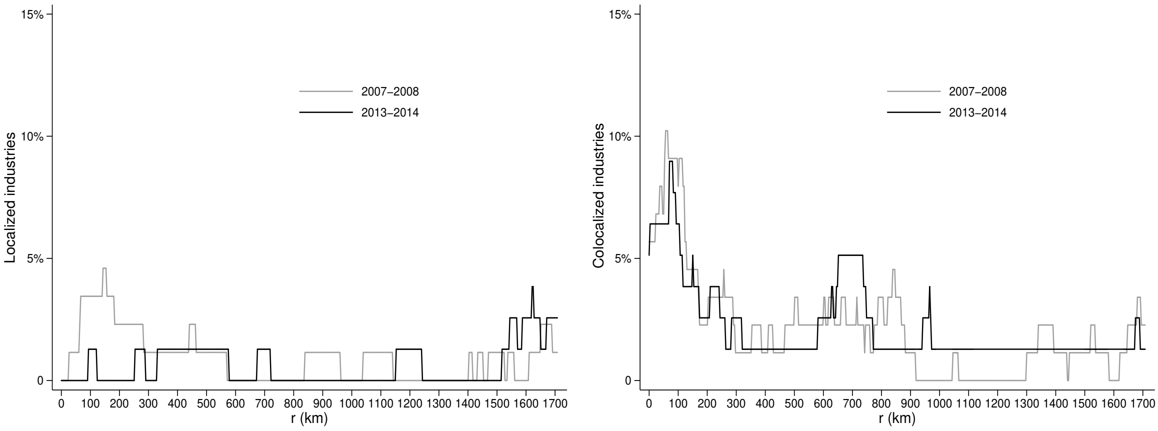
Figure 2. K-density estimates for manuf. of pharmaceutical and food products

Table 1. Localization and colocalization of employment-weighted new establishments

	Localization				Colocalization			
	2007-2008		2013-2014		2007-2008		2013-2014	
	# of ind.	%	# of ind.	%	# of ind.	%	# of ind.	%
Localized	13	14.94	10	12.82	38	42.53	21	26.92
Dispersed	7	8.05	8	10.26	19	21.84	21	26.92
Random	67	77.01	60	76.92	31	35.63	36	46.15
	87[a]	100	78[b]	100	87	100	78	100

Notes: After the restriction imposed (minimum of 10 plants in each sector): [a] 16 industries were dropped and [b] 15 industries were dropped. Source: Prepared by the author based on estimates.

Similarly, but looking at the general colocation patterns, 42.53% and 26.92% of employment-weighted entrants are colocalized with existing establishments in the same periods, as can be seen on the left panel of Table 1. These results indicate that the new manufacturing establishments in Brazil tend to be more colocalized than in the UK (9% as shown by Duranton and Overman, 2008), which may indicate that agglomeration forces, especially those associated with specialization, are more important in Brazil due to the different urban structures. In contrast, 21.84% and 26.92% are codispersed. Furthermore, unlike the localization results, our evidence for colocalization indicates that entrants are colocalized mainly at short distances, as can be seen in Figure 3 (b). For instance, among the colocalized industries for which entrants are closer to existing establishments are *manufacture of pharmaceutical products* - CNAE 212 and *manufacture of other food products* - CNAE 109 (as shown in Figures 2 (b) and (d); other food industries such as *fruit & vegetable canning* - CNAE 103, *manufacture of starch products* - CNAE 106, and *manufacture of furniture* - CNAE 310 (less than 10 km); *prepress and graphic finishing services* - CNAE 182 (around 40 km); *manufacture of wood products* - CNAE 162 and *finishing of textile articles* - CNAE 134 (around 70 km).[7]



(a) Significant localization by distance

(b) Significant colocalization by distance

Figure 3. Shares of industries in which entrants are localized and colocalized

Our evidence so far provides a details about location and colocation patterns of entrants in Brazilian manufacturing activities, but it does provide any information about the associated agglomeration forces. Note, for example, that there a tendency exists both in 2007-2008 and 2013-2014 for colocalization to take place at short distances (less than 70 km). This is consistent with the idea that the local economy, in particular the presence of existing establishments in the same industry, can determine the location choice of new establishments. In a more detailed analysis, similar to that conducted by Klier and McMillen (2008) for the automotive industry in the US, but looking at all manufacturing activities and focusing on the spatial scope of agglomeration economies, in the next sections we explore the importance of proximity to existing establishments to the location choice of entrants.

3 Empirical strategy

3.1 Model specification

Our objective is to evaluate the spatial extent of agglomeration economies on the quantity of birth of new establishments and the employment levels that they choose in a given geographical area. To do so, it is necessary to solve two important secondary issues: (i) choosing the appropriate empirical model for our problem; and (ii) determining the geographical areas of analysis. The first has an immediate solution given the characteristic of our problem, count data. Thus, we model our problem using count models, more specifically the Poisson model, widely used in the literature (see Arauzo-Carod et al., 2010 for a survey) and directly related to the problem described in the previous section from the establishment's location choice problem using the random profit maximization approach (see Carlton, 1983). For the second point, our finely geocoded employer-employee data allow us to freely define spatial units of measurement. All Brazilian territory is divided into around 8.5 million cells exogenously determined with one square kilometer

[7]On the other hand, the maximum of the distribution of the codispersed industries occurs around 1,080 km.

each (1 km × 1 km).[8] The definition of such small microgeographic areas helps us to deal with a common problem in studies about agglomeration economies, sorting. Technically, if we evaluate the birth of firms in period t and the local characteristics in $t-1$, there is no simultaneity (Jofre-Monseny et al., 2014), but firms can rank the eligible locations one year earlier, i.e., spatial sorting. However, our exogenously defined set of cells is outside the firms' choice. For example, it is difficult to think that the choice of the city of a new establishment is random. Furthermore, within the city itself, the districts can still be chosen. On the other hand, the new firm does not choose the specific cell because this depends on the availability of land and land use.

Once these initial conditions are established, we assume that new establishments are opened at locations chosen from among of the square kilometer (cell $z = 1, \cdots, Z$) of Brazilian territory. Since our spatial unit of measurement is homogeneous, additional concerns regarding differences in the sizes of the geographical units are not necessary.[9] To capture the spatial extent of the agglomeration economies, we construct five concentric rings: 0-1, 1-5, 5-10, 10-20, and 20-40 km from the centroid of each cell, to measure our agglomeration variables. These variables are measured as usually done in location choice studies (see, e.g., Figueiredo et al., 2002; Jofre-Monseny et al., 2014; Li et al., 2020, just to cite a few), i.e., the own-industry j employment, emp_{jrt}, and the employment in other industries, emp_{-jrt}, in ring r and period t. The first measure captures the local intra-sectoral externalities (localization economies) and is associated with proximity to existing employment in establishments belonging to the same industry as the new establishment. The spatial concentration of plants of a particular manufacturing sector operates as a pool of favorable conditions, providing local specialized labor, sharing of intermediate input markets and generating knowledge spillovers. The second measure captures more general inter-sectoral local externalities associated with concentration of general economic activity in a particular area.[10] This type of externality can be internalized by all plants in the same area. Thus, for each 3-digit industry, the following Poisson model is estimated:

$$\mathbf{E}(\text{Y}_{jzt+2}) = \exp\left(\sum_{r} \beta_{jr}^{loc}\text{emp}_{jrt} + \sum_{r} \beta_{jr}^{urb}\text{emp}_{-jrt} + \mathbf{X}_{jzt}\tau + \gamma_d\right) \tag{1}$$

where Y_{jzt+2} is the number of new plants or the number of jobs in these new plants in industry j, cell z and period $t+2$,[11] $\mathbf{X}_{jzt}$ is a vector of control variables containing location determinants other than agglomeration economies, and γ_d is the district fixed effect. We use the Poisson pseudo-maximum likelihood (PPML) estimator with multiple high-dimensional fixed effects recently developed by Correia et al. (2020) to deal with the large number of district fixed effects.

The main concern in the estimation of equation 1 is to obtain unbiased and consistent estimates for the set of parameters of interest, β_{jr}^{loc}, where $r = 1, \cdots, 5$. By including the fixed district effect, we control for any observed and unobserved characteristics fixed in time and specific to the district, which minimizes possible biases of omitted variables. Furthermore, we also test the robustness of the estimates by including a comprehensive set at cell and municipality level control variables that may affect the new establishments' location choice. The next subsection presents more details of these control variables.

3.2 Control variables

To address the omitted variable bias concerns, we test the robustness of our estimates by including control variables for economic environment, previously existing transportation infrastructure, geographical

[8]Recently, with the availability of microgeographic data, other studies have used similar strategies (see, e.g., Larsson, 2014; Andersson et al., 2014; 2019; Li et al., 2020). Obviously, most of these cells do not have any kind of economic activity. Technically, an obvious criterion to select our study's geographic area is that most cells are uninhabited areas, such as forests, lakes and rivers.

[9]Rosenthal and Strange (2003), for example, used Zip code areas of the US to deflate both births and new-establishment employment.

[10]Since we use only employment in manufacturing, these variables measure part of the urbanization economies.

[11]We calculate our outcomes variables in the periods 2007-2008, 2009-2010, 2011-2012 and 2013-2014 and the explanatory variables in 2006, 2008, 2010 and 2012. This is a common strategy in studies like this (see, e.g., Rosenthal and Strange; 2003; Jofre-Monseny, 2009).

characteristics, and local development policies around the place chosen by the new establishment.

We begin with the economic environment around a specific cell. There is abundant evidence that incumbent local industrial structures can influence the level of local entrepreneurship. In particular, the presence of many small establishments is associated with employment growth in start-ups (see, e.g., Chinitz, 1961; Glaeser and Kerr, 2009; Rosenthal and Strange, 2010; Ghani et al., 2014). Thus, as in Rosenthal and Strange (2003) and Li et al. (2020), we include proxies for local industry organization and industry diversity. To be more precise, we include two Herfindahl indices within 40 km of the cell's centroid z. The first index captures, for example, the competition effects around the z and is measured for each 3-digit industry by $\sum_j (\text{emp}_{ijzt}/\text{emp}_{jzt})^2$, where emp_{ijzt} is the employment level of plant i in industry j in the region within 40 km of z in period t, and emp_{jzt} is the employment level of industry j in the region within 40 km of z in period t. This variable controls for local industrial organization around z. The second index captures the local diversity of economic activities and is measured by $\sum_j (\text{emp}_{jzt}/\text{emp}_{zt})^2$, where $\text{emp}_{jzt}/\text{emp}_{zt}$ is the industry j's share of total employment within 40 km of the centroid of z in period t.

Also at the cell level, we include a set of time-invariant transport and geographic controls. There is a broad set of evidence that transportation infrastructure can influence the location choice and productivity of plants (see, e.g., Holl, 2004a; 2004b; 2016; Mayer and Trevien, 2017; Gibbons et al., 2019), thus including the set of transportation controls eliminates the effects of the previously installed transportation infrastructure on our estimates. At the same time, the proximity of rivers or lakes can affect the choice of location of resource-intensive industries by reducing input transportation costs (Ellison and Glaeser, 1999; Ellison et al., 2010; Rosenthal and Strange, 2001). Furthermore, by including these variables, we also control for zoning and planning restrictions, something relevant in an intra-urban context. The transportation controls include the distance between each cell's centroid and the nearest airports, public ports, railways, federal highways, and state highways. The geographic control is the distance between each cell's centroid and the nearest river. We interact these time-invariant controls with time effects to capture differential trends across cells.

Local development policies can also play an important role in attracting new investments (e.g., Glaeser et al., 2010b; Chatterji et al., 2014), and thus affect the location choice of new enterprises. So, at municipal level, we include values of capital expenditures (investments) and housing and town planning expenses per hundred thousand inhabitants, as a proxy for the quality of previously existing urban infrastructure; municipal taxes per hundred thousand inhabitants to control for differences in tax costs between municipalities; tax incentive policies implemented previously by local governments; and exports and imports per hundred thousand inhabitants, as a proxy for the access of firms previously located in the municipality to the international market. We also include homicides per hundred thousand inhabitants, which provides an indicator of the efficiency of public security policies implemented by local governments; and traffic fatalities per hundred thousand inhabitants, which acts as an indicator of the quality of the municipal public transportation system.

3.3 Remaining heterogeneities and control function approach

As discussed earlier, we use district fixed effects and a comprehensive set of control variables to minimize the bias in the estimation of localization economy effects. Even so, we cannot guarantee that the problem of omitted variables is completely solved. Various factors can make the estimates of parameters associated with these variables biased upward or downward. For example, one potential source of endogeneity is the factors in $\mathsf{A}(\mathbf{y})$, in particular, any local unobserved characteristics that can affect productivity can cause higher existing business concentrations, while at the same time, also attract a higher number of new establishments (Combes et al., 2008; Li et al., 2020).

To address these potential concerns, we complement the analysis with a shift-share instrumental variable that exploits the changes in national employment growth specific to the industry (a "shift") to generate exogenous variation at the concentric ring level, in a control function approach.[12] The measure consists of the growth in employment that would have occurred had each industry in a concentric ring grown at

[12] This is an approach to estimate nonlinear models with endogenous explanatory variables (Terza et al., 2008; Wooldridge, 2014). To illustrate the application of this approach to this study, we follow Navarro (2008) and Cameron and Trivedi (2013).

its national rate of growth (Bartik, 1991). More specifically, we use 33 industries at the 4-digit level to calculate the instrument for employment growth of eight 3-digt industries colocalized at short distances (mentioned in subsection 2.2) in each concentric ring r at time t. Formally:

$$\text{IV}_{jrt} = \sum_k \sum_c \omega_{rc} \text{emp}_{kc1995} \ln\left(\frac{\text{emp}_{kt} - \text{emp}_{kRt}}{\text{emp}_{k1995} - \text{emp}_{kR1995}}\right), \text{ with } \omega_{rc} = \frac{\mathsf{A}_{r\cap c}}{\mathsf{A}_c} \tag{2}$$

where $\mathsf{A}_{r\cap c}$ is the intersection area between concentric ring r and municipality c; A_c is total area of the municipality; emp_{kc1995} is the employment in industry k at the 4-digit level belonging to industry j at the 3-digit level in municipality c in reference year;[13] emp_{kt} is the national employment in industry k and year $t = (1999, \cdots, 2003)$; emp_{kRt} is the employment in industry k in area $R = \sum_r \mathsf{A}_r$ and year t. In other words, like Moretti and Thulin (2013), we are discounting from national employment in industry k the sum of employment in the five concentric rings in industry k.

This instrument isolates the variation that comes from nationwide changes at the 4-digit industry level k and uses the sum of the industrial mix components to calculate the variation of 3-digit industry j. To understand the logic, consider as example two concentric rings with the same size and the same share of manufacturing jobs in 1995, but a different industry mix within 3-digit manufacturing. If employment in a given industry increases nationally (where we remove from nationwide changes the local changes within a radius of 40 km), the concentric ring where industry employs a larger share of the labor force experiences a positive shock to the labor demand in the manufacturing sector. On the other hand, if employment in a given industry decreases, the concentric ring experiences a negative shock to the labor demand in the manufacturing sector (Moretti, 2010).

Using this instrument to construct a control function, the coefficients can be estimated in two steps. In the first step, our proxies for localization economies in each concentric ring, emp_{jrt}, is regressed using ordinary least squares (OLS) estimation on observed characteristics, presented in the previous subsection, and the instruments. In the second step, the count model is estimated with the residuals of the first step entering as a control for the unobserved confounder bias. This procedure is also known in the literature as two-stage residual inclusion (2SRI) and is more appropriate to deal with endogenous explanatory variables in nonlinear models than the extension of the popular linear two-stage least squares estimator for nonlinear models (Terza et al., 2008; Terza, 2017). While this approach is widely employed in health econometric research (see, e.g., Stuart et al., 2009; Lazuka, 2018; Ghanbariamin and Chung, 2020), its application in econometric studies of regional and urban economics is rare, despite the widespread use of nonlinear models with potentially endogenous explanatory variables, especially in studies of locational choice.

To illustrate this procedure, consider the following version of the count model presented in equation 1:

$$\mathbf{E}(\text{Y}_{jzt+2}) = \exp\left(\sum_r \beta^{loc}_{jr} \text{emp}_{jrt} + \sum_r \beta^{urb}_{jr} \text{emp}_{-jrt} + \mathbf{X}_{jzt}\tau + \mathbf{u}_{jzt}\right) \tag{3}$$

where $\mathbf{u}_{jzt}$ are the effects not captured by the control variables included. Thus, the first step equations are given by:

$$\text{emp}_{jrt} = \sum_r \delta_{jr} \text{IV}_{jrt} + \sum_r \theta^{urb}_{jr} \text{emp}_{-jrt} + \mathbf{X}_{jzt}\lambda + \mathbf{w}_{jrt}, \; r = 1, \cdots, 5. \tag{4}$$

where IV_{jrt} is the shift-share instrumental variable, and $\mathbf{w}_{jrt}$ are omitted factors that can influence local industry concentration. If the terms $\mathbf{u}_{jzt}$ and $\mathbf{w}_{jrt}$ are correlated for any of the reasons mentioned above, then emp_{jrt} and $\mathbf{u}_{jzt}$ are correlated, so the Poisson regression of $\mathbf{E}(\text{Y}_{jzt+2})$ on emp_{jrt} and the other co-variables yields inconsistent parameter estimates (Cameron and Trivedi, 2013). However, we can obtain consistent estimates if $\mathbf{u}_{jzt} = \rho_j \mathbf{w}_{jrt} + \varepsilon_{jzt}$ with ε_{jzt} independent of $\mathbf{w}_{jrt}$, and estimating Poisson regression by substituting the term $\mathbf{u}_{jzt}$ in equation 3 by $\mathbf{w}_{jrt}$ estimated from the first step by OLS (Wooldridge,

[13] We defined the reference year as the first year (1995) for which the National Classification of Economic Activities - CNAE is available. Thus, we used the first version of the CNAE (or CNAE 1.0), which contains 564 four-digit groups of industries, of which 268 are manufacturing industries. We merged the old CNAE version with the new version CNAE 2.0 (available from 2006) from the correspondence tables provided by IBGE, available at https://cnae.ibge.gov.br/.

1997; 2010). That is, the second step equation is:

$$\mathbf{E}(\mathrm{Y}_{jzt+2}) = \exp\left(\sum_r \beta_{jr}^{loc}\mathrm{emp}_{jrt} + \sum_r \beta_{jr}^{urb}\mathrm{emp}_{-jrt} + \mathbf{X}_{jzt}\tau + \rho_j \hat{\mathbf{w}}_{jrt}\right) \tag{5}$$

Additional concerns are related to the estimates obtained in the second step. We use $\hat{\mathbf{w}}_{jrt}$ as opposed to $\mathbf{w}_{jrt}$, i.e., a generated regressor, so we need to adjust the standard error estimates in the second step to take this extra source of variation into account (Petrin and Train, 2010; Cameron and Trivedi, 2013). We implemented bootstrapping to adjust the standard errors of the second step.[14] The estimated coefficient for $\hat{\mathbf{w}}_{jrt}$ provides the direction of unobserved confounder bias, and its statistical significance will indicate whether the variable emp_{jrt} is indeed endogenous (Wooldridge, 1997).

4 Results

4.1 Baseline results

We showed in subsection 2.2 that the new establishments in some industries are colocalized at short distances. To be more precise, we showed that the entrants engaged in *manufacture of pharmaceutical products* - CNAE 212, *manufacture of other food products* - CNAE 109, *fruit & vegetable canning* - CNAE 103, *manufacture of starch products* - CNAE 106, *manufacture of furniture* - CNAE 310, *prepress and graphic finishing services* - CNAE 182, *manufacture of wood products* - CNAE 162, and *finishing of textile articles* - CNAE 134 are colocalized less than 70 km. Now, we evaluate if entrants' location choice in these industries can be affected by agglomeration economies. In particular, we evaluate whether the probability of an entrant being located in a specific cell depends on proximity to existing establishments in the same industry. Tables 2 and 3 present estimates of equation 1 when our outcome variables are the number of new establishments per cell and new-establishment employment per cell, respectively.

For example, consider initially the first estimated coefficient for the *prepress and graphic finishing services* - CNAE 182 (column 6 of Table 2), for which the localization effects are among the most pronounced. Adding 100 prepress workers up to 1 km would generate, on average, an increase of 37.6% in the expected number of births and 35.7% in the expected number of employees (as can be seen in the top panel of Tables 2 and 3). Adding 100 additional employees to the 1-5 km ring would result, on average, an increase of 5% in the expected number of births. On the other hand, when the outcome is new-establishment employment, the coefficient is not significant. For distances larger than 10 km, the coefficients are not significant, except for the 20-40 km ring in Table 3, which is negative, suggesting some kind of competition at greater distances. Note that these estimates are free of any bias caused by omitted variables that are time-invariant at the district level and any specific tendency related to previously existing transportation infrastructure, geographic characteristics, and local development policies.

We also provide the results of the models without district fixed effects and/or without control variables for the eight industries and note the dispersion in our data. Based on the descriptive statistics by industry, in general there is overdispersion in the data for new-establishment employment, so the negative binomial model is more appropriate. When the district fixed effects and/or control variables or urbanization variables are omitted from the Poisson or negative binomial models, there is, of course, a reduction in the magnitude of the estimated coefficients, and in most cases only the coefficients associated with the smaller rings remain positive and strongly significant for both firm birth and new-establishment employment. However, the pattern of attenuation with distance remains.[15]

Looking in particular at entrants engaged in *manufacture of pharmaceutical products* - CNAE 212 (column 1 of Tables 2 and 3), the results also indicate that the probability of an entrant choosing a cell, but not the new-establishment employment, is higher when there are already other establishments in the same industry up to 1 km from location of the new establishment. In contrast, existing establishments

[14] We performed 400 bootstrap replications following the examples in Cameron and Trivedi (2013).

[15] Due to space limitations, this information can be made available upon request.

Table 2. Spatial scope of localization and urbanization externalities - plant birth. Poisson regression

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Pharma. Products CNAE 212	Food Products CNAE 109	Fruit & vegetable Canning CNAE 103	Starch Products CNAE 106	Furniture CNAE 310	Prepress Services CNAE 182	Wood Products CNAE 162	Fin. of textile Articles CNAE 134
Localization Effects								
0 to 1 km	9.83e-04**	4.70e-04***	1.55e-03**	1.94e-03***	6.54e-04***	3.76e-03***	2.20e-03***	1.91e-03***
	(3.90e-04)	(5.22e-05)	(6.53e-04)	(4.13e-04)	(7.29e-05)	(4.34e-04)	(2.83e-04)	(5.25e-04)
1 to 5 km	7.62e-05	1.79e-04***	5.64e-04	-2.32e-05	1.10e-04**	5.05e-04**	9.64e-05	3.05e-04
	(3.29e-04)	(3.80e-05)	(5.18e-04)	(3.75e-04)	(5.00e-05)	(2.00e-04)	(2.11e-04)	(2.11e-04)
5 to 10 km	1.18e-04	8.64e-05***	4.46e-05	-5.99e-05	-6.20e-06	-2.32e-04	1.13e-04	-4.25e-04**
	(2.09e-04)	(3.32e-05)	(4.84e-04)	(3.16e-04)	(5.03e-05)	(1.78e-04)	(1.78e-04)	(1.72e-04)
10 to 20 km	1.32e-05	4.47e-05	1.69e-04	-8.02e-05	-5.95e-05	8.49e-05	-8.96e-06	-1.40e-06
	(1.78e-04)	(2.96e-05)	(4.09e-04)	(2.64e-04)	(3.73e-05)	(1.51e-04)	(1.52e-04)	(1.11e-04)
20 to 40 km	1.04e-04	-1.10e-05	4.25e-04	1.82e-04	3.02e-06	-1.99e-05	-1.18e-04	-1.89e-04**
	(1.87e-04)	(2.64e-05)	(3.31e-04)	(2.34e-04)	(2.56e-05)	(1.70e-04)	(1.00e-04)	(9.47e-05)
Urbanization Effects								
0 to 1 km	1.95e-04***	1.32e-04***	1.52e-04***	2.02e-04***	1.52e-04***	7.04e-05***	1.77e-04***	1.84e-04***
	(4.21e-05)	(8.47e-06)	(3.38e-05)	(2.33e-05)	(8.56e-06)	(1.37e-05)	(1.41e-05)	(1.64e-05)
1 to 5 km	3.72e-05**	1.89e-05***	1.12e-05	1.89e-05**	2.12e-05***	2.33e-05***	1.01e-05*	2.44e-05***
	(1.83e-05)	(3.13e-06)	(9.87e-06)	(8.84e-06)	(2.89e-06)	(5.06e-06)	(5.35e-06)	(5.87e-06)
5 to 10 km	-2.73e-06	-6.44e-06***	2.55e-06	-3.20e-06	4.12e-06*	-3.53e-06	-2.35e-06	-7.39e-09
	(1.19e-05)	(2.19e-06)	(6.57e-06)	(6.88e-06)	(2.18e-06)	(3.78e-06)	(4.24e-06)	(4.10e-06)
10 to 20 km	-3.36e-06	-4.27e-06**	-4.13e-06	8.23e-06*	7.38e-07	-2.44e-06	1.24e-07	-3.15e-06
	(9.24e-06)	(1.72e-06)	(5.56e-06)	(4.32e-06)	(1.44e-06)	(3.22e-06)	(2.57e-06)	(2.49e-06)
20 to 40 km	1.13e-06	-2.06e-06	-4.46e-06	3.55e-06	2.67e-06**	-5.06e-06*	-1.23e-06	2.62e-06
	(4.90e-06)	(1.39e-06)	(4.34e-06)	(3.94e-06)	(1.16e-06)	(2.93e-06)	(2.29e-06)	(2.14e-06)
Average Change in Localization Effect per km								
0.5 to 3 km	-3.63e-04	-1.16e-04	-3.93e-04	-7.87e-04	-2.18e-04	-1.30e-03	-8.40e-04	-6.43e-04
3 to 7.5 km	9.32e-06	-2.06e-05	-1.15e-04	-8.14e-06	-2.58e-05	-1.64e-04	3.67e-06	-1.62e-04
7.5 to 15 km	-1.40e-05	-5.56e-06	1.65e-05	-2.72e-06	-7.11e-06	4.23e-05	-1.62e-05	5.65e-05
15 to 30 km	6.06e-06	-3.71e-06	1.71e-05	1.75e-05	4.17e-06	-6.98e-06	-7.30e-06	-1.25e-05
# of district FE	81	1874	381	663	1579	395	957	426
Pseudo R^2	0.1482	0.0743	0.1178	0.1245	0.1011	0.1257	0.1151	0.1606
Pseudo-LL	-523.8067	-26,834.97	-2,694.916	-4,507.093	-24,581.67	-4,507.879	-9,027.774	-5,030.079
Observations	19,151	118,376	47,495	61,350	111,668	56,812	82,976	58,389

Notes: This table reports the localization and urbanization effects when the dependent variable is the number of new establishments in each cell. Heteroscedasticity-robust standard errors are reported in parentheses. All columns include the diversification and competition control variables, transport and geographic controls, municipality level controls, and district fixed effects. The transportation controls include the distance to the nearest airport, public port, railway, federal highway, and state highway interacted with time effects. The geographic control is the distance to the nearest river interacted with time effects. The municipality level controls include proxies for insertion in international trade (exports and imports), municipal taxes, capital investments, housing and town planning expenses, homicides and traffic fatalities. Change per kilometer is computed by differencing the adjacent localization coefficients and dividing by the number of kilometers between the midpoints. Significance level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Source: Prepared by the author based on estimates.

located at larger distances, for example, 5, 10, 20 and 40 km, do not affect the probability of an entrant's choice. This pattern confirms our initial impressions (as shown in Figures 1 (a) and 2 (b)) that localization economies can determine the location choice of new establishments in this industry. More than that, there is a pattern of spatial decay of the effects generated by localization externalities consistent with the idea of gains generated by spillovers at short distances. In addition, estimates for the more general effects, associated with urbanization externalities, are also highly concentrated, much stronger up to 1 km, five times smaller up to 5 km and not significant thereafter. This indicates that the effects generated by proximity to existing establishments in other industries are also attenuated with distance.

Similarly, column 2 of Tables 2 and 3 report the results for the entrants engaged in *manufacture of other food products* - CNAE 109. In general, the results point in the same direction, i.e., there is a pattern of spatial decay in both localization and urbanization effects, but unlike what is observed for pharmaceutical products, in the food industry the localization externalities extend to large distances, precisely up to 10 km for firm birth outcome and 20 km for new-establishment employment, consistent with the pattern observed previously in the Figures 1 (b) and 2 (d). This is a low-tech industry localized (at large distances) relative to manufacturing activity as a whole, a pattern consistent with the evidence found here indicating that this large spatial extension of localization effects is associated with sharing the local labor market. Therefore, although our model identifies agglomeration effects based on within-district variation of the data, our results are broadly consistent with previous works that was based on between-city variation in the data, i.e., also providing evidence of Marshallian agglomeration forces that act at larger spatial scales as labor market pooling (Rosenthal and Strange, 2003; 2020). In subsection 4.3 we provide a more detailed comparison of our results with those obtained for other countries. Note also that the effects of urbanization are attenuated up to 5 km, which suggests that the presence of existing establishments in other industries at short distances around the cell can act as an attraction force, increasing the probability of choosing the specific cell.

Estimates in columns 3-8 of both Tables 2 and 3 for all other industries previously classified as colocalized at short distances also indicate that the effects of both proximity to establishments in the same

Table 3. Spatial scope of localization and urbanization externalities - new employment. Poisson regression

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Pharma. Products CNAE 212	Food Products CNAE 109	Fruit & vegetable Canning CNAE 103	Starch Products CNAE 106	Furniture CNAE 310	Prepress Services CNAE 182	Wood Products CNAE 162	Fin. of textile Articles CNAE 134
				Localization Effects				
0 to 1 km	4.79e-04	1.54e-03***	4.19e-03***	3.12e-03***	7.93e-04***	3.58e-03**	5.78e-03***	2.08e-03*
	(9.98e-04)	(2.92e-04)	(1.23e-03)	(1.06e-03)	(2.02e-04)	(1.49e-03)	(8.63e-04)	(1.22e-03)
1 to 5 km	-9.69e-04	4.55e-04**	1.58e-03	-1.42e-03	-8.70e-05	8.32e-04	1.57e-03**	-4.25e-05
	(8.58e-04)	(2.17e-04)	(1.39e-03)	(9.60e-04)	(1.70e-04)	(5.77e-04)	(6.75e-04)	(4.21e-04)
5 to 10 km	-5.45e-05	3.49e-04*	1.78e-03	-1.62e-03	-3.94e-05	-8.41e-04	5.17e-04	-7.10e-04*
	(6.41e-04)	(1.84e-04)	(1.15e-03)	(1.06e-03)	(1.25e-04)	(6.00e-04)	(7.18e-04)	(3.97e-04)
10 to 20 km	-2.38e-04	2.44e-04*	1.38e-03	-1.54e-03*	-1.75e-05	-1.56e-04	-7.17e-04	-1.53e-04
	(8.07e-04)	(1.45e-04)	(1.08e-03)	(8.69e-04)	(9.06e-05)	(3.73e-04)	(6.04e-04)	(2.31e-04)
20 to 40 km	-2.49e-04	5.47e-05	-2.16e-04	-3.22e-04	6.20e-05	-1.13e-03*	-6.68e-04*	-5.48e-05
	(1.03e-03)	(1.22e-04)	(1.23e-03)	(6.82e-04)	(1.14e-04)	(6.04e-04)	(3.97e-04)	(1.66e-04)
				Urbanization Effects				
0 to 1 km	1.97e-04**	1.93e-04***	2.22e-04***	3.22e-04***	1.99e-04***	4.70e-05**	1.91e-04***	1.94e-04***
	(9.28e-05)	(3.19e-05)	(6.46e-05)	(6.62e-05)	(1.85e-05)	(2.18e-05)	(2.81e-05)	(5.84e-05)
1 to 5 km	4.48e-05	9.68e-06	2.28e-05	7.64e-05***	2.12e-05**	5.94e-06	-4.33e-06	4.24e-05***
	(5.08e-05)	(1.10e-05)	(3.70e-05)	(2.10e-05)	(9.79e-06)	(1.35e-05)	(1.62e-05)	(1.50e-05)
5 to 10 km	-2.38e-05	-1.84e-05*	-4.85e-06	2.46e-05**	-5.39e-07	-8.57e-07	2.25e-05	1.23e-05
	(4.06e-05)	(1.07e-05)	(1.59e-05)	(1.23e-05)	(6.73e-06)	(1.04e-05)	(2.03e-05)	(9.65e-06)
10 to 20 km	2.21e-05	-1.38e-05	-3.67e-05***	-7.51e-07	-1.33e-06	-3.91e-06	1.84e-05**	-1.99e-05***
	(2.20e-05)	(9.76e-06)	(1.20e-05)	(1.52e-05)	(3.86e-06)	(9.67e-06)	(7.83e-06)	(4.89e-06)
20 to 40 km	9.71e-06	-3.90e-06	-3.76e-05***	1.82e-05	3.70e-07	-2.08e-06	2.34e-05**	1.58e-06
	(1.84e-05)	(7.69e-06)	(1.14e-05)	(1.17e-05)	(4.27e-06)	(6.98e-06)	(9.14e-06)	(5.64e-06)
			Average Change in Localization Effect per km					
0.5 to 3 km	-5.79e-04	-4.34e-04	-1.04e-03	-1.82e-03	-3.52e-04	-1.10e-03	-1.68e-03	-8.50e-04
3 to 7.5 km	2.03e-04	-2.35e-05	4.46e-05	-4.34e-05	1.06e-05	-3.72e-04	-2.33e-04	-1.48e-04
7.5 to 15 km	-2.45e-05	-1.40e-05	-5.36e-05	1.02e-05	2.93e-06	9.14e-05	-1.65e-04	7.43e-05
15 to 30 km	-7.47e-07	-1.26e-05	-1.06e-04	8.12e-05	5.30e-06	-6.52e-05	3.25e-06	6.54e-06
# of district FE	41	1374	214	385	1,175	245	682	308
Pseudo R^2	0.5028	0.2535	0.5626	0.3411	0.2558	0.2908	0.4039	0.3095
Pseudo-LL	-24,369.76	-113,055.7	-14,575.2	-32,034.53	-71,729.64	-12,520.69	-37,538.29	-16,137.08
Observations	11,089	105,114	33,052	42,262	98,583	42,243	66,745	46,925

Notes: This table reports the localization and urbanization effects when the dependent variable is the number of new employments in each cell. Heteroscedasticity-robust standard errors are reported in parentheses. All columns include the diversification and competition control variables, transport and geographic controls, municipality level controls, and district fixed effects. The transportation controls include the distance to the nearest airport, public port, railway, federal highway, and state highway interacted with time effects. The geographic control is the distance to the nearest river interacted with time effects. The municipality level controls include proxies for insertion in international trade (exports and imports), municipal taxes, capital investments, housing and town planning expenses, homicides and traffic fatalities. Change per kilometer is computed by differencing the adjacent localization coefficients and dividing by the number of kilometers between the midpoints. Significance level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Source: Prepared by the author based on estimates.

industry and to establishments in other industries attenuate rapidly with distance. Observe also that in all cases the localization effects are more important than urbanization effects. In particular, for the first concentric ring of employment, the coefficient of the localization employment variable is, on average, 13.6 times larger than the coefficient of the corresponding urbanization employment variable when the outcome is the birth of firms and 20 times larger when the outcome is the new-establishment employment. For the second concentric ring, the difference is, on average, 12.1 times larger when the outcome is the birth of firms and not significant when the outcome is the new-establishment employment. As explained by Rosenthal and Strange (2003), this provides a clear distinction between urbanization and localization economies, since it is expected that the gains from information spillovers and the ability to share both intermediate inputs and specialized labor diminish monotonically with increasing distance, while the urbanization effects can be of any sign because of tradeoff between the benefits and congestion costs of locating near densely developed areas. This evidence also points in the same direction as the results found by Henderson (1986) for Brazil and US, Nakamura (1985) for Japan with aggregate data, and more recently, based on within-city variation, by Li et al. (2020) for China, but is more robust since we control for several heterogeneities not included in the previous studies.

In summary, our key geographical results are that for most industries, the localization economies attenuate rapidly in the first few kilometers. Another way to observe this pattern, as in Rosenthal and Strange (2003), is presented at the bottom of the tables, i.e., change per kilometer (CPkm) in the localization effects for each industry, measured by the difference in coefficients between each of the adjacent pairs of concentric rings divided by the number of kilometers between the midpoints of the two rings. For births (Table 2), the averaging across all eight industries of the ratio of CPkm values in the 0.5 km to 3 km range relative to the 3 km to 7.5 km range, $\mathrm{CPkm}_{(0.5 \text{ to } 3)}/\mathrm{CPkm}_{(3 \text{ to } 7.5)}$, is 9.5, while the ratio $\mathrm{CPkm}_{(3 \text{ to } 7.5)}/\mathrm{CPkm}_{(7.5 \text{ to } 15)}$ is -6.93, and the ratio $\mathrm{CPkm}_{(7.5 \text{ to } 15)}/\mathrm{CPkm}_{(15 \text{ to } 30)}$ is 4.88. When looking at the new-establishment employment (Table 3), the analogous values are 13.97, 7.22, and 0.88 respectively.

4.2 Control function results

In this subsection we report the results of the estimates in two steps as a robustness test, i.e., when we include in addition to the comprehensive set of control variables presented above, the fitted residual from the first-stage regression. In Panel A (top) of Table 4, we present the estimated coefficients in the second step when our outcome is the count of new establishments in each cell. In Panel B (bottom), we present the results when our outcome is the level of new-establishment employment by cell.

The endogeneity of emp_{jrt} ($r = 1, \cdots, 5$) can be tested based on the coefficient of the first-stage residual. For most industries we reject the null hypothesis of exogeneity of emp_{j1t}, i.e., the exogeneity of own-industry employment in the first concentric ring. For the concentric rings farthest from the cell's centroid, in most cases we fail to reject the null hypothesis of exogeneity of emp_{jrt}.

The coefficient of own-industry employment in the first concentric ring (0-1 km) is now, on average, 4.7 and 11.7 times as large as under the exogeneity assumption when the outcome is the birth of firms and new-establishment employment, respectively. For the second concentric ring (1-5 km), the statistically significant coefficients in both cases (Tables 2 and 4) are now, on average, 4.8 times large for the outcome firm birth, while there is no major change when the outcome is employment in new establishments. For comparison, consider again the *prepress and graphic finishing services* - CNAE 182. Adding 100 prepress workers up to 1 km would generate, on average, an increase of 60% in the expected number of births as can be seen in the column 6 of Panel A of Table 4, while the equivalent coefficient when the outcome is the new-establishment employment is not significant (Panel B). The coefficients of own-industry employment in the second concentric ring are not significant.

Table 4. Second stage of two-step estimates of localization effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Pharma. Products CNAE 212	Food Products CNAE 109	Fruit & vegetable Canning CNAE 103	Starch Products CNAE 106	Furniture CNAE 310	Prepress Services CNAE 182	Wood Products CNAE 162	Fin. of textile Artifacts CNAE 134
Panel A: The dependent variable is births of new establishments								
0 to 1 km	9.02e-03**	1.14e-03***	6.40e-03**	5.33e-03***	3.86e-03***	6.02e-03**	7.60e-03***	1.63e-02***
	(3.50e-03)	(2.96e-04)	(2.57e-03)	(1.18e-03)	(5.14e-04)	(3.01e-03)	(8.31e-04)	(5.16e-03)
1 to 5 km	-4.53e-04	2.90e-04***	2.31e-04	4.88e-04	-5.99e-05	8.65e-05	5.45e-04***	2.16e-03***
	(3.91e-04)	(5.13e-05)	(6.59e-04)	(2.99e-04)	(6.49e-05)	(3.92e-04)	(1.66e-04)	(5.23e-04)
5 to 10 km	1.76e-04	-1.31e-05	-1.48e-04	-6.02e-05	2.40e-04***	3.55e-05	5.14e-04**	-6.02e-04*
	(1.77e-04)	(3.07e-05)	(4.53e-04)	(2.69e-04)	(6.07e-05)	(2.40e-04)	(2.07e-04)	(3.42e-04)
10 to 20 km	7.10e-05	1.34e-05	5.05e-04***	1.64e-04	7.59e-06	1.92e-04*	-1.72e-04	8.59e-04***
	(8.98e-05)	(1.48e-05)	(1.95e-04)	(1.20e-04)	(2.54e-05)	(1.12e-04)	(1.06e-04)	(9.71e-05)
20 to 40 km	-3.94e-05	-4.48e-05***	2.18e-04	1.31e-04	6.05e-05***	-4.16e-04***	1.89e-04***	-2.57e-05
	(9.56e-05)	(1.16e-05)	(1.38e-04)	(1.01e-04)	(7.57e-06)	(9.75e-05)	(4.65e-05)	(6.00e-05)
Panel B: The dependent variable is new-establishment employment								
0 to 1 km	2.80e-02*	2.81e-03***	2.20e-02***	4.55e-03*	6.19e-03***	6.40e-03	1.01e-02***	3.27e-02***
	(1.59e-02)	(9.23e-04)	(7.48e-03)	(2.51e-03)	(9.74e-04)	(1.38e-02)	(2.56e-03)	(9.07e-03)
1 to 5 km	-1.90e-03	4.78e-04**	-3.84e-03	2.32e-03***	-3.05e-04**	8.67e-04	-6.92e-04	3.26e-03***
	(1.77e-03)	(2.24e-04)	(2.79e-03)	(7.92e-04)	(1.53e-04)	(1.44e-03)	(6.84e-04)	(1.03e-03)
5 to 10 km	7.55e-04	-1.75e-04	3.70e-04	-2.51e-04	5.44e-04***	8.64e-04	1.84e-03*	-1.77e-03**
	(7.02e-04)	(2.30e-04)	(2.04e-03)	(5.15e-04)	(2.03e-04)	(1.27e-03)	(9.74e-04)	(7.23e-04)
10 to 20 km	4.74e-04	2.15e-04	1.47e-03**	7.71e-04**	-7.68e-06	3.22e-04	-5.15e-04	1.12e-03***
	(3.57e-04)	(1.34e-04)	(6.83e-04)	(3.06e-04)	(8.36e-05)	(4.13e-04)	(3.65e-04)	(2.87e-04)
20 to 40 km	3.51e-04*	-1.47e-04	-1.88e-04	2.85e-04	7.02e-05***	-9.20e-04**	4.52e-04***	-9.03e-05
	(2.11e-04)	(1.04e-04)	(6.25e-04)	(1.99e-04)	(2.70e-05)	(4.49e-04)	(1.30e-04)	(1.17e-04)

Notes: This table reports the localization effects when we include a control function to address endogeneity concerns (equation 5). All models are estimated with 139,527 observations. All columns report the results from Poisson regressions where the dependent variable is the births of new establishments (Panel A) and the new-establishment employment (Panel B), and the variable of interest is the number of workers in the same industry in each concentric ring. All columns include the urbanization, diversification and competition variables, along with the transport and geographic controls, as presented in the previous section. Standard errors based on 400 bootstrap replications are reported in parentheses. Significance level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Source: Prepared by the author based on estimates.

Another interesting example, for which the difference is more pronounced, is the *manufacture of pharmaceutical products* - CNAE 212. Under the exogeneity assumption, adding 100 pharmaceutical products workers up to 1 km would generate, on average, an increase of 9.8% in the expected number of births, but is not significant for employment at new establishments (as reported in the column 1 of Tables 2 and 3), while in the two-stage estimates, adding the same number of workers up to 1 km would generate, on average, an increase of 90% in the expected number of births and 280% in the expected number of employees (column 1 of Panels A and B of Table 4).

While these comparisons between the coefficients estimated under the exogeneity assumption and the two-stage estimated coefficients provide some indications about the sign and magnitude of the bias, the main result of this subsection is that the pattern of attenuation with distance can also be clearly observed in the two-stage results. Although in both cases the second stage results show more specific patterns. One is in column 8 for *finishing of textile articles* - CNAE 134, where a negative effect appears at 10 km and then becomes positive at 20 km (Panels A and B); another is in column 5 for *manufacture of furniture* - CNAE 310, where a negative effect appears at 5 km and then becomes positive at 10 km (Panel B). This may represent some kind of competition generated by the emergence of new clusters. But note that in both cases the residuals of the first-stage equation are not significant and we do not reject the null hypothesis of exogeneity. In summary, our main findings remain valid, indicating that our results are robust to possible biases caused by the potential endogeneity of our main explanatory variables.

Our results for Brazil are consistent with theoretical models of urban areas and previous empirical evidence for other countries. For example, Rosenthal and Strange (2003) found similar results for US (localization effects attenuated around 10 km). For developing countries, the evidence on the subject is scarcer. Indeed, there is only the recent study of Li et al. (2020), who found that the localization effects in some industries are attenuated more rapidly with distance in China than in developed countries. As in that study, when we compare our general results with Rosenthal and Strange (2003)'s results, we find that localization effects are also attenuated more rapidly in Brazil. Note that in most cases for both births and new-establishment employment, the positive effects disappear after 5 km, which may suggest that localization effects are attenuated more rapidly because of the local urban infrastructure. We present a more detailed analysis of this phenomenon in the next subsection.

Also exploring the spatial scope of agglomeration economies, but using wages as an outcome, Rosenthal and Strange (2008) found that agglomeration economies attenuate rapidly with distance in the US. In particular, the human capital spillovers were found to be stronger up to 8 km from the individual's workplace. For other developed countries, also using wages as an outcome variable, Di Addario and Patacchini (2008) for the Italy found that urbanization externalities are attenuated up to 12 km, and more recently Håkansson and Isacsson (2019) found that urbanization externalities are attenuated around 25 km in Sweden. These studies, although addressing a different question than ours, also present evidence in the same direction, i.e., the agglomeration forces are heterogeneous within cities, so the use of geographically aggregated data does not allow exploring in detail the spatial scope of these externalities.

4.3 Other key industries and comparison with US and Chinese results

We have thus far presented the spatial scope of agglomeration economies in Brazil. In this subsection, we contrast our estimates with estimates from Rosenthal and Strange (2003) for the US and Li et al. (2020) for China to reveal possible differences in attenuation patterns with distance between countries. Rosenthal and Strange (2003) estimated the determinants of firm birth for six industries (software - SIC 7371, 7372, 7373, and 7375, food products - SIC 20, apparel - SIC 23, printing and publishing - SIC 27, fabricated metals - SIC 34, and industrial and commercial machinery - SIC 35). Li et al. (2020) estimated the determinants of firm birth for all Chinese manufacturing industries. Except for the software industry, we contrast the US and Chinese results for the remaining five industries with the results obtained for similar industries in Brazil. In the previous sections we have already presented the results for *manufacture of other food products* - CNAE 109 and *prepress and graphic finishing services* - CNAE 182, so to make comparison possible, we complement the analysis here by including *manufacture of wearing apparel* - CNAE 141, *manufacture of metal structures* - CNAE 251 and *manufacture of machinery* - CNAE 282.[16]

Before we begin a comparison by industry, we highlight some important general differences between the previous two papers and the present study. Unlike Rosenthal and Strange (2003) and Li et al. (2020), we have panel data, which allows us to control for any observed and unobserved heterogeneities fixed in time in different areas within cities (districts). We use an exogenous microgeographic spatial partitioning structure that, as we discussed earlier, minimizes sorting bias. We also use a comprehensive set of control variables for previously existing transportation infrastructure and geographic features around the cell and local development policies.

Now, to get a detailed view of localization effects on the firm birth by industry, each panel in Figure 4 reports the results of the comparison for a specific industry (results from Tables 2). As in Li et al. (2020),

[16]Due to space limitations, this results can be made available upon request.

to allow for easy comparison and interpretation, here we also define the vertical axis in each figure so that the magnitude of the spillover effects in the machinery industry within the first ring in the corresponding study is equal to one and all other spillover effects are measured relative to this value. The horizontal axis measures the spatial distance between firms in the same industry. Since the scale and the measurement unit vary among the three studies,[17] we first convert the unit of measurement used by Rosenthal and Strange (2003) (miles) to kilometers, and then we obtain the intermediate values of the estimated coefficients by linear interpolation.

Some interesting patterns emerge from these comparisons. We start with the most general. After adjusting the different scales and distance units used, for 3 of the 5 industries analyzed (machinery, printing and food products), the attenuation of localization economies is faster in Brazil than in the US (as can be seen in Figure 4 (a-c)). For 2 industries (machinery and apparel), the attenuation patterns in Brazil are similar to those in China. The contrast can imply that the knowledge spillovers and/or labor pooling mechanism in the US, and to a lesser degree in China, acts over larger distances, so that firms farther from each other can still share knowledge and the specialized labor market. This evidence conforms very well with the pattern of high concentration of manufacturing industries in Brazil.

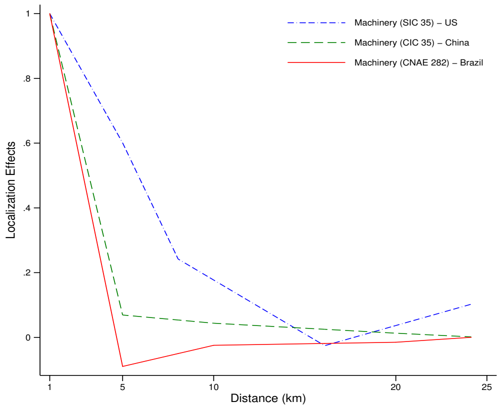
In the machinery industry, the attenuation in Brazil is similar to China, while in printing, the attenuation is faster in Brazil. For the food industry in Figure 4 (c), the attenuation is faster in Brazil relative to the US but not to China. This contrast with the US may be related to the different extension of the transportation infrastructure in the two countries. Similar arguments were also used by Li et al. (2020) to explain the differences in attenuation patterns that are slower in the US than in China, and also apply very well to the characteristics of the costly transportation system in Brazil. A costly and underdeveloped transportation system that makes collaboration between more distant firms hard may restrict the effects of location economies to short distances. Relative to the results for China, the slower attenuation of localization effects in the food industry in Brazil may be associated with the supply of inputs spread throughout all regions of the country. For the metal industry, depicted in Figure 4 (d), the attenuation is slower in Brazil, both in relation to the US and China.

Another interesting pattern that contrasts with what has been discussed so far can be seen in Figure 4 (e) for the apparel industry. The attenuation of localization economies in this industry in the US is very fast when compared to the pattern observed in the corresponding industry in Brazil. For example, for this industry in the US, the localization effects disappear (are not significant) at distances greater than 1.6 km, while in Brazil for both births and new-establishment employment these positive effects disappear at distances greater than 5 km. This pattern is consistent with the argument that the apparel industry in the US could be more directly engaged in designing and advertising, which benefits more from knowledge spillovers because it requires more idea sharing and networking (Li et al.; 2020). The apparel industry in Brazil, as in China, is more associated with manufacturing processing, which depend less on knowledge spillovers. In Figure 4 (f), we plot the estimated coefficients excluding results for the US, because they have different scales, to make it easier to compare the results for Brazil with those for China. The spatial decay of the effects is similar to that observed in China.

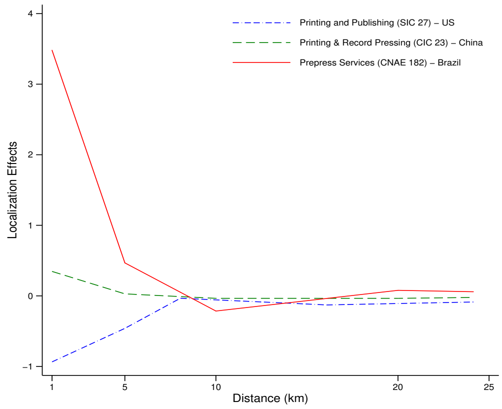
5 Concluding remarks

The objective of the article has been to examine the spatial scope of agglomeration economies in Brazil. In order to do so, we use a unique and rich microgeographic database for all Brazilian manufacturing industries and estimate the local determinants of the number of births per square kilometer and their associated employment levels as functions of the own-industry employment in different distance bands, controlling for the economic environmental characteristics around the site chosen by the new establishment. To better understand the geographical distribution of the new establishments, initially we address location and colocation patterns of new manufacturing plants using the nonparametric approach of Duranton and

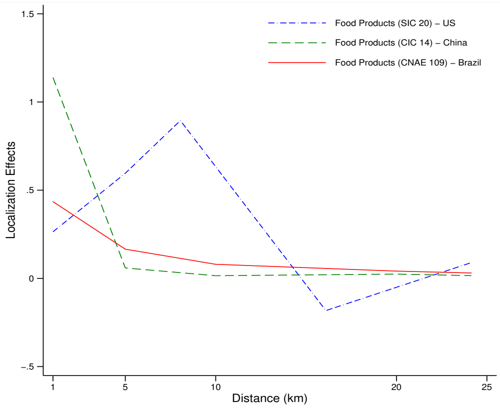
[17]For example, Rosenthal and Strange (2003) used four concentric rings set between 0-1 miles (0-1.6 km), 1-5 miles (1.6-8 km), 5-10 miles (8-16 km), and 10-15 miles (16-24 km). In turn, Li et al. (2020) used five concentric rings set between 0-1 km, 1-5 km, 5-10 km, 10-20 km, and 20-30 km. For comparison purposes, we set the maximum distance on the horizontal axis equal to 24 km (distance of the largest ring used by Rosenthal and Strange (2003)).



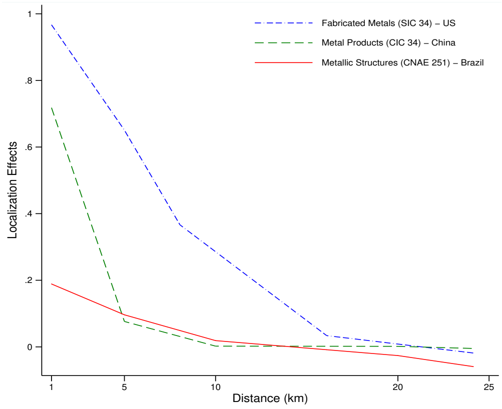
(a) Machinery



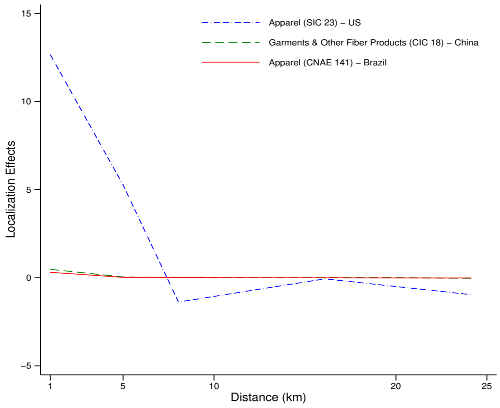
(b) Printing



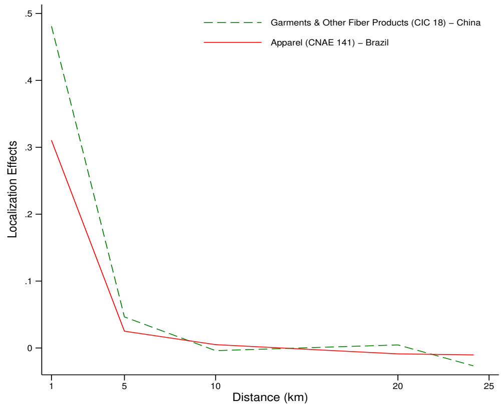
(c) Food Products



(d) Metal Products



(e) Apparel



(f) Apparel without US

Figure 4. Attenuation of localization economies for five selected manufacturing industries in comparison with Rosenthal and Strange (2003) and Li et al. (2020)

Overman (2005; 2008). After obtaining the pattern of geographic distribution of the new plants, we analyze the spatial scope of the agglomeration economies considering mainly the industries classified as colocalized at short distances.

Unlike previous studies on the same subject, such as Rosenthal and Strange (2003) and Li et al. (2020), we use panel data, which allows us to control for time-fixed observed and unobserved heterogeneities at the district level and a comprehensive set of control variables for the economic environment, previously installed transportation infrastructure, geographical characteristics, and local development policies applied to the place chosen by the new establishment. We also use instrumental variables to address any remaining sources of heterogeneities in a control function approach.

From our initial nonparametric investigation, two main features emerge revealing details not available in the literature for developing countries, although they were found by Duranton and Overman (2008) for the UK. Among all manufacturing activities, 14.94% and 12.82% of entrants in 2007-2008 and 2013-2104 were localized while 77% in both periods did not have statistically different location patterns from those observed for existing establishments. The other feature is that, in contrast to what is observed for localization, there is a colocation pattern at short distances, which suggests that agglomeration economies are important. Furthermore, 42.53% and 26.92% of entrants were colocalized with existing establishments in the same periods.

In our parametric investigation, we find that in nearly all cases for both births and new-establishment employment, localization effects (own-industry employment at 3-digit level) are important, mainly up to 5 km from the birthplace. These results show that the infrastructure of cities in Brazil and the way manufacturing activities are geographically distributed benefit the emergence of new establishments in areas where there is existing geographical concentration. Moreover, both the births of new establishments and the level of employment they choose are higher in places that present high local specialization, which consequently affects the local choice of the new plant. This pattern conforms very well with the high geographic concentration of manufacturing observed in Brazil and with the high interregional mobility of workers. This evidence is also consistent with previous results for other countries indicating that localization effects attenuate rapidly over the first few kilometers. Our results are robust to the inclusion of a comprehensive set of controls, district fixed effects and also the inclusion of instrumental variables in a control function approach to deal with the potential endogeneity of our key explanatory variables.

As mentioned earlier, this paper provides evidence about a topic relevant to the formulation of economically efficient public policies focused on manufacturing entrepreneurship that considers the intrinsic forces of attraction in the market. In summary, although it is not in the scope of this study to determine the sources of agglomeration that cause the observed pattern, in general our evidence is in accordance with the spatial scope of action of the three Marshallian agglomeration forces, namely knowledge spillovers at short distances, labor market pooling and shared inputs at greater distances. Exploring these sources in detail is on the agenda for future studies.

References

Andersson, M., Klaesson, J., and Larsson, J. P. (2014). How local are spatial density externalities? neighbourhood effects in agglomeration economies. *Regional Studies*, 50(6):1082–1095.

Andersson, M., Larsson, J. P., and Wernberg, J. (2019). The economic microgeography of diversity and specialization externalities-firm-level evidence from swedish cities. *Research Policy*, 48(6):1385–1398.

Arauzo-Carod, J.-M., Liviano-Solis, D., and Manjón-Antolín, M. (2010). Empirical studies in industrial location: an assessment of their methods and results. *Journal of Regional Science*, 50(3):685–711.

Bartik, T. J. (1991). Who benefits from state and local economic development policies?

Barufi, A. M. B., Haddad, E. A., and Nijkamp, P. (2016). Industrial scope of agglomeration economies in Brazil. *The Annals of Regional Science*, 56(3):707–755.

Cameron, A. C. and Trivedi, P. K. (2013). *Regression analysis of count data*, volume 53. Cambridge UP.

Carlton, D. W. (1983). The location and employment choices of new firms: An econometric model with discrete and continuous endogenous variables. *The Review of Economics and Statistics*, pages 440–449.

Chatterji, A., Glaeser, E., and Kerr, W. (2014). Clusters of entrepreneurship and innovation. *Innovation Policy and the Economy*, 14(1):129–166.

Chauvin, J. P., Glaeser, E., Ma, Y., and Tobio, K. (2017). What is different about urbanization in rich and poor countries? cities in Brazil, China, India and the United States. *Journal of Urban Economics*.

Chinitz, B. (1961). Contrasts in agglomeration: New York and pittsburgh. *American Economic Review*.

Combes, P.-P., Duranton, G., and Gobillon, L. (2008). Spatial wage disparities: Sorting matters! *Journal of Urban Economics*, 63(2):723–742.

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

Di Addario, S. and Patacchini, E. (2008). Wages and the city. evidence from Italy. *Labour Economics*.

Duranton, G. (2015). Growing through cities in developing countries. *World Bank R.*, 39(1):39–73.

Duranton, G. and Overman, H. G. (2005). Testing for localization using micro-geographic data. *The Review of Economic Studies*, 72(4):1077–1106.

Duranton, G. and Overman, H. G. (2008). Exploring the detailed location patterns of uk manufacturing industries using microgeographic data. *Journal of Regional Science*, 48(1):213–243.

Ellison, G. and Glaeser, E. L. (1997). Geographic concentration in us manufacturing industries: a dartboard approach. *Journal of Political Economy*, 105(5):889–927.

Ellison, G. and Glaeser, E. L. (1999). The geographic concentration of industry: does natural advantage explain agglomeration? *American Economic Review*, 89(2):311–316.

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

Faberman, R. J. (2011). The relationship between the establishment age distribution and urban growth. *Journal of Regional Science*, 51(3):450–470.

Figueiredo, O., Guimaraes, P., and Woodward, D. (2002). Home-field advantage: location decisions of portuguese entrepreneurs. *Journal of Urban Economics*, 52(2):341–361.

Fritsch, M. and Wyrwich, M. (2016). The effect of entrepreneurship on economic development—an empirical analysis using regional entrepreneurship culture. *Journal of Economic Geography*, 17(1):157–189.

Fu, S. (2007). Smart café cities: Testing human capital externalities in the Boston metropolitan area. *Journal of Urban Economics*, 61(1):86–111.

Ghanbariamin, R. and Chung, B. W. (2020). The effect of the national kidney registry on the kidney-exchange market. *Journal of Health Economics*, 70:102301.

Ghani, E., Kerr, W. R., and O'connell, S. (2014). Spatial determinants of entrepreneurship in india. *Regional Studies*, 48(6):1071–1089.

Gibbons, S., Lyytikäinen, T., Overman, H. G., and Sanchis-Guarner, R. (2019). New road infrastructure: the effects on firms. *Journal of Urban Economics*, 110:35–50.

Glaeser, E. L. (2007). Entrepreneurship and the city. Technical report, NBER.

Glaeser, E. L., Kerr, S. P., and Kerr, W. R. (2015). Entrepreneurship and urban growth: An empirical assessment with historical mines. *Review of Economics and Statistics*, 97(2):498–520.

Glaeser, E. L. and Kerr, W. R. (2009). Local industrial conditions and entrepreneurship: how much of the spatial distribution can we explain? *Journal of Economics & Management Strategy*, 18(3):623–663.

Glaeser, E. L., Kerr, W. R., and Ponzetto, G. A. (2010a). Clusters of entrepreneurship. *Journal of Urban Economics*, 67(1):150–168.

Glaeser, E. L., Rosenthal, S. S., and Strange, W. C. (2010b). Urban economics and entrepreneurship. *Journal of Urban Economics*, 67(1):1–14.

Håkansson, J. and Isacsson, G. (2019). The spatial extent of agglomeration economies across the wage earnings distribution. *Journal of Regional Science*, 59(2):281–301.

Hansen, E. R. (1987). Industrial location choice in sao paulo, brazil: a nested logit model. *Regional science and Urban economics*, 17(1):89–108.

Head, K., Ries, J., and Swenson, D. (1995). Agglomeration benefits and location choice: Evidence from Japanese manufacturing investments in the united states. *Journal of International Economics*.

Henderson, J. V. (1986). Efficiency of resource usage and city size. *Journal of Urban Economics*.

Holl, A. (2004a). Manufacturing location and impacts of road transport infrastructure: empirical evidence from spain. *Regional Science and Urban Economics*, 34(3):341–363.

Holl, A. (2004b). Transport infrastructure, agglomeration economies, and firm birth: empirical evidence from portugal. *Journal of Regional Science*, 44(4):693–712.

Holl, A. (2016). Highways and productivity in manufacturing firms. *Journal of Urban Economics*.

Jofre-Monseny, J. (2009). The scope of agglomeration economies: Evidence from Catalonia. *Papers in Regional Science*, 88(3):575–590.

Jofre-Monseny, J., Marín-López, R., and Viladecans-M., E. (2014). The determinants of localization and urbanization economies: Evidence from the location of new firms in Spain. *Journal of Regional Science*.

Klier, T. and McMillen, D. P. (2008). Evolving agglomeration in the US auto supplier industry. *Journal of Regional Science*, 48(1):245–267.

Larsson, J. P. (2014). The neighborhood or the region? reassessing the density–wage relationship using

geocoded data. *The Annals of Regional Science*, 52(2):367–384.
Lautert, V. and Araújo, N. C. M. d. (2007). Concentração industrial no Brasil no período 1996-2001: uma análise por meio do índice de ellison e glaeser (1994). *Economia Aplicada*, 11(3):347–368.
Lazuka, V. (2018). The long-term health benefits of receiving treatment from qualified midwives at birth. *Journal of Development Economics*, 133:415–433.
Leff, N. H. (1972). Desenvolvimento econômico e desigualdade regional: origens do caso brasileiro. *Revista Brasileira de Economia*, 26(1):3–22.
Li, J., Li, L., and Liu, S. (2020). Attenuation of agglomeration economies: Evidence from the universe of chinese manufacturing firms.
Marshall, A. (1890). *Principles of Economics*. Macmillan London.
Mayer, T. and Trevien, C. (2017). The impact of urban public transportation evidence from the paris region. *Journal of Urban Economics*, 102:1–21.
Moretti, E. (2010). Local multipliers. *American Economic Review*, 100(2):373–77.
Moretti, E. and Thulin, P. (2013). Local multipliers and human capital in the united states and sweden. *Industrial and Corporate Change*, 22(1):339–362.
Nakamura, R. (1985). Agglomeration economies in urban manufacturing industries: a case of japanese cities. *Journal of Urban Economics*, 17(1):108–124.
Navarro, S. (2008). Control functions. In *The New Palgrave Dictionary of Economics*. Palgrave M. UK.
Paes, N. L. and Siqueira, M. L. (2005). Análise dos efeitos econômicos da implantação do princípio do destino na cobrança do ICMS e suas implicações sobre a pobreza e a desigualdade de renda. *Revista EconomiA*, 6(3):91–126.
Paes, N. L. and Siqueira, M. L. (2008). Desenvolvimento regional e federalismo fiscal no Brasil: em busca da igualdade na distribuição de receitas. *Economia Aplicada*, 12(4):707–742.
Petrin, A. and Train, K. (2010). A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research*, 47(1):3–13.
Rocha, R. d. M., Araújo, J. E. S., and de Almeida, E. T. (2019). As indústrias da tranformação são concentradas geograficamente? um teste empírico para o Brasil (2002-2014). *Nova Economia*.
Rosenthal, S. S. and Strange, W. C. (2001). The determinants of agglomeration. *Journal of Urban Economics*, 50(2):191–229.
Rosenthal, S. S. and Strange, W. C. (2003). Geography, industrial organization, and agglomeration. *Review of Economics and Statistics*, 85(2):377–393.
Rosenthal, S. S. and Strange, W. C. (2008). The attenuation of human capital spillovers. *Journal of Urban Economics*, 64(2):373–389.
Rosenthal, S. S. and Strange, W. C. (2010). Small establishments/big effects: Agglomeration, industrial organization and entrepreneurship. In *Agglomeration Economics*, pages 277–302. UCP.
Rosenthal, S. S. and Strange, W. C. (2020). How close is close? the spatial reach of agglomeration economies. *Journal of Economic Perspectives*, 34(3):27–49.
Silveira Neto, R. d. M. (2005). Concentração industrial regional, especialização geográfica e geografia econômica: evidências para o Brasil no período 1950-2000. *Revista Econômica do Nordeste*.
Stuart, B. C., Doshi, J. A., and Terza, J. V. (2009). Assessing the impact of drug use on hospital costs. *Health Services Research*, 44(1):128–144.
Tatsch, A. L., Ruffoni, J., Batisti, V. d. S., and Roxo, L. A. T. (2015). Análise de políticas para aglomerações no Brasil e em países europeus selecionados. *Planejamento e Políticas Públicas*.
Terza, J. V. (2017). Two-stage residual inclusion estimation in health services research and health economics. *Health Services Research*, 53(3):1890–1899.
Terza, J. V., Basu, A., and Rathouz, P. J. (2008). Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *Journal of Health Economics*, 27(3):531–543.
Varsano, R. (1997). A guerra fiscal do ICMS: quem ganha e quem perde. *TD 500 - IPEA*.
Wooldridge, J. M. (1997). Quasi-likelihood methods for count data. *Handbook of Applied Econometrics*.
Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.
Wooldridge, J. M. (2014). Quasi-maximum likelihood estimation and testing for nonlinear models with endogenous explanatory variables. *Journal of Econometrics*, 182(1):226–234.