

Buczowska, Sabina; de Lapparent, Matthieu

Conference Paper

Location choices of newly created establishments: spatial patterns at the aggregate level

54th Congress of the European Regional Science Association: "Regional development & globalisation: Best practices", 26-29 August 2014, St. Petersburg, Russia

Provided in Cooperation with:

European Regional Science Association (ERSA)

Suggested Citation: Buczowska, Sabina; de Lapparent, Matthieu (2014) : Location choices of newly created establishments: spatial patterns at the aggregate level, 54th Congress of the European Regional Science Association: "Regional development & globalisation: Best practices", 26-29 August 2014, St. Petersburg, Russia, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/124422>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Location choices of newly created establishments: spatial patterns at the aggregate level

Sabina Buczkowska^{1a}, Matthieu de Lapparent^a

^a *University Paris-Est, French Institute of Science and Technology for Transport, Development and Networks, 14-20 Boulevard Newton, F-77447 Marne la Vallée Cedex 2, France*

Abstract

This paper explores the problems associated with the location choice of newly created establishments at the aggregate level. Much work has been done in this domain, however, several issues arise when analyzing involved phenomena, which scholars have yet to fully explore: 1) addressing the excess of zeros problem in the location choice model in highly heterogeneous geographic areas and 2) determining an appropriate way to accommodate spatial effects for location decisions. We tested models that include both stocks of pre-existing establishments and variables that represent measures of accessibility to the workforce and population, proximity to shops, services, transport infrastructure, availability of land, as well as prices and tax levels. We concluded that an establishment does not act in isolation during its decision-making processes and that it is likely to be influenced by other establishments located nearby. When selecting the appropriate location in which to set up in the market, an establishment may consider not only the characteristics of a particular area, but also the characteristics of neighboring zones. Having estimated 84 nested and non-nested count data models, we found that the hurdle models are preferred for taking into account the presence of excess zeros. Hurdle models offer greater flexibility in modeling zero outcomes and relax the assumption that the zero observations and the positive observations come from the same data generating process. In addition, the paper finds that the models tested with the distance matrix indicate that the incorporation of spatial spillovers leads to an enhancement in the models' performance.

Keywords: Location choice model; Count data models; Hurdle model; Spatial spillovers

JEL Classification: C35, D21, R12

1. Introduction

The present paper explores the problems associated with the location choice of newly created establishments at the aggregate level within the Paris metropolitan area in the year 2007. Applications of disaggregate discrete choice modeling to analysis of birth, death, evolution, and location of establishments in the Paris region have been carried out by de Palma et al. (2008) and Motamedi (2008). We also refer the reader to the European SustainCity Project². Much work has been done in this domain. However, several issues

¹Corresponding author. Tel.: +33 0181668614. *E-mail address:* sabina.buczkowska@ifsttar.fr (S. Buczkowska).

²See <http://www.sustaincity.org/publications>.

arise when analyzing involved phenomena, which scholars have yet to fully explore, most notably the excess of zeros problem in the location choice model in highly heterogeneous geographic areas and determining an appropriate way to incorporate spatial effects into the model. For example, spatial effects can be divided into two categories: the first considers the interactions between establishments that do not act in isolation and are thus likely to be influenced by the decisions of other players in the market, while the second looks at correlations across alternative choices. Inspired by spatial econometrics techniques, e.g., Anselin, (1988), Lambert et al. (2010), Klier and McMillen (2008), we seek solutions to these and related problems.

The proposed approach is descriptive, aggregate, and partial. For the purposes of the present research, all data are gathered at the municipality level. In the first step, we estimate standard non-spatial nested and non-nested count data models. In the second step, we run these count data models that accommodate spatial spillovers for location decisions. We compare all the results obtained and choose the best performing hurdle model, after which a more thorough interpretation of its results is undertaken.

The need to specify the model becomes apparent when dealing with various types of location choice models, which range from those focusing on employment location estimated at the job level to those which focus on establishment or firm location and the associated decision-making processes. Implementing the latter form of location choice models, we acknowledge that the decision to open a new unit is made at the establishment or firm level. We define an establishment as a distinct economic unit that produces goods or services at a single physical location. In contrast, a firm is a legal entity that consists of one or more establishments or plants under common ownership and control (van Wissen, 2000). Thus, it is a newly created establishment that offers job opportunities to members of a given population provided they possess the appropriate skills and desired characteristics. For this reason, the unit of analysis on which we focus is that of the establishment.

The rest of the paper is organized as follows. In Section 2, we review the literature on both location choice models and the initial attempts to incorporate spatial effects into such models. In Section 3, we describe the area of study, which is the Paris region. In Section 4, we describe data, presenting statistical sources and their harmonization in order to carry out the empirical application. Descriptive statistics are presented in Section 5. We give a brief description of the count data models and develop our parametric statistical model in Section 6, starting from the logit-Poisson model, a type of hurdle-Poisson model, and discuss how to extend it in order to meet the needs of the present research. In Section 7, the results from our research are discussed. In the final section, we conclude and elaborate on extensions of the proposed approach.

2. Review of empirical literature

2.1. Location models

A number of empirical studies suggest variables that play a significant role in influencing the establishment's location choice. We briefly describe those which are most relevant. According to Arauzo-Carod (2005) and Maoh (2005), an establishment will decide to enter a market if it detects a potential business opportunity

and if the capital and human resources are available. An establishment must choose a type of activity that reflects both its technological and organizational level and its minimum size. With these parameters in mind, the establishment makes a decision as to where to locate. Since choosing a particular area in which to locate is critical to the establishment's success or failure (Strotmann, 2007), the main task of the establishment is to identify the places which offer the highest possible potential profit. Local economic climate in terms of its ability to attract new entrants, as well as the structure of establishments at the local level are other factors that can influence the decision (Liviano-Solís and Arauzo-Carod, 2011). Locations with lower input costs are also more likely to experience a higher number of births of new establishments (Rocha, 2008).

Sutton (2007) notices that ranking sectors according to the degree of (entry-exit) turbulence appears to offer similar results across countries, suggesting that there are some sector-specific factors at work in molding this pattern. Rocha (2008) and Liviano-Solís and Arauzo-Carod (2011) concur, finding that local employment density attracts new entrants in related sectors and has a positive impact on the establishment's productivity. However, when this density is too high, the effect becomes negative due to congestion costs, including for example, high land prices and costly commuting. This results in an inverted U-shape profile between the effect of concentration of economic activity and site attractiveness. The relationship is initially positive, but becomes negative after a certain threshold is crossed. Maoh and Kanaroglou (2005, 2007) state that agglomeration effects tend to be more significant in particular activity sectors, such as retail and services.

Local authorities try to offer different incentives in a bid to attract establishments to less developed areas and accelerate the development of the whole region. These include, but are not limited to, supply of infrastructure, designation of new building zones, and tax reductions. Empirical results from various studies undertaken in different countries worldwide demonstrate that these policies may have positive, negative, or neutral effects on the location choice and activity distribution.

Special local fiscal arrangements are common practice for attracting establishments and have received considerable attention from economists (e.g., Bondonio and Greenbaum, 2007). Offers of subsidies and tax reductions are popular yet controversial tools. Low taxes are another common instrument to attract establishments. Tax reductions show positive effects on activity development. Switzerland is a prime example (Bodenmann, 2011). Devereux et al. (2007) show that in the UK the impact of subsidies is rather small on the location choice but they are more effective in dense areas when a large number of establishments is already present. Bondonio and Greenbaum (2007) show that in the US enterprise zone programs (EZ thereafter), which hope to convince establishments to locate in less developed zones by offering them subsidies, have a significant positive impact on the creation of new establishments. However, they also show that EZ policies tend to increase business closures. Neumark and Kolko (2010) show that the overall impact of the EZ program in California is ineffective and causes no positive impact on the number of jobs.

Providing special infrastructure is an effective but costly way to raise the attractiveness of a region. De Bok (2004) shows that accessibility has a positive effect on business services. Bodenmann and Axhausen (2012) claim that this also holds for other sectors and demonstrate that interventions regarding transport

infrastructure have a larger impact in the densely populated areas of Switzerland. Devereux et al. (2007) show that the effects of transport infrastructure projects are unequally distributed spatially. All actions have larger impacts in denser regions. Siebert (2000) states that increasing accessibility has a relatively small effect on regional development.

Bodenmann and Axhausen (2010) summarize how the significance of location factors has changed over time. In the 1990s, the most critical location factors were human capital, agglomeration effects, and accessibility. Currently, a positive economic climate and direct costs (earning levels, tax burdens, etc.) play larger roles. More curious readers are encouraged to acquaint themselves with the referenced literature and the work of Arauzo-Carod et al. (2010). They provide a review of over fifty papers on location choice modeling with a focus on the analyses of location decisions of new industrial establishments or firms using appropriate econometric models. They describe the establishment/firm location determinants, the econometric methods used in these investigations, and their principal results³.

2.2. First attempts to incorporate spatial effects in location choice models

As underscored by Nguyen et al. (2012), an establishment does not act in isolation during its decision-making process and is likely to be influenced by other establishments located nearby. When choosing an appropriate place in which to set up in the market, an establishment can take into account not only the characteristics of a particular area but also those of its surroundings. The reason for doing so is the spatial dependence of neighboring areas. In addition, the degree of spatial correlations is expected to be greater among choice alternatives (the options available to the decision-maker) that are close to one another. Jayet (2001) proves the existence of interactions among units located in space and demonstrates that their intensity decreases with distance. Thus, two types of spatial effects may be considered: 1) interactions between establishments and 2) dependencies among alternatives. Spatial effects can be incorporated in location choice models when modeling the observable explanatory variables and the unobservable components. However, most often, these spatial effects are either not properly treated or are completely ignored in the analysis of

³The most commonly used establishment/firm location determinants and the signs of their estimates used in both discrete choice and count data models according to the review of Arauzo-Carod et al. (2010) are: agglomeration economies (+,-: positive or negative effect), previous entries in the own sector (+), existing plants (+), own-industry employment (+), sectoral diversity (+,-), sectoral specialization (+,-), market size (+), #competitors (+), #small and medium enterprises (+), establishment/firm size (+), productivity (+), unemployment (+,-), industrial employment share (+), services employment share (+), business services (+), share of employees in R&D (+), human capital (+,-), knowledge spillovers (+), skilled workforce (+), education (-), schooling (+), existence of high schools (+), overall R&D investment (+), R&D facilities (+), high-ranking hotels (+), population density (+,-), distance to urban areas (-), land area (+,-), land costs (-), entry costs (-), taxes (-), corporate tax rate (+,-), taxes on labor (-), labor costs (+,-), wages (+,-), income per capita (+), purchasing power per inhabitant (+), GDP (+), poverty (-), local demand (+), supplier accessibility (+), government spendings (+), promotional subsidies (+), labor and capital subsidies (+), economic promotion (+), investment climate (+), infrastructure (+), transportation infrastructure (+), road infrastructure (+), distance to highway (-), rail infrastructure (-), airports facilities (+), travel time to airport (-), energy costs (+,-), and environmental regulation (-).

the establishment's location. An example of the former can be seen in Bhat and Guo (2004) who model the spatial dependence using the mixed logit model, with an arbitrary spatial allocation instead of basing these spatial effects on a systematic specification. Furthermore, even if spatial effects are present they are not incorporated in traditional discrete choice models. For instance, multinomial logit models (MNL) are commonly used in location choice processes, yet they are based on several simplifying assumptions, such as independent and identical Gumbel distribution (i.i.d.) of random components of the utilities and the absence of heterogeneity and autocorrelation in the model. These simplifying assumptions limit the ability of the model to represent the true structure of the choice process (Mohammadian et al., 2005).

Traditional discrete choice modeling methods are often based on the assumption of independence among choice alternatives, which according to Sener et al. (2011) is not appropriate. Furthermore, Sener et al. (2011) claim that the estimations of the parameters of the standard logit models are biased and inconsistent. Instead, they suggest using the generalized spatially correlated logit (GSCL) to model the residential location choice behavior and they incorporate the correlations between the alternatives. Their model can be enhanced to accommodate random taste variations across decision-makers. The authors compare the results obtained with their GSCL model against those obtained using the standard multinomial logit and the spatially correlated logit (for which only correlations between the adjacent alternatives are considered). Moreover, they do not limit themselves to a few choice alternatives since GSCL models eliminate the need for any kind of simulation and can be estimated by direct maximum likelihood techniques. They criticize prior studies that were carried out with a limited number of alternatives in the choice set. Miyamoto et al. (2004) estimate a mixed logit model with the error autocorrelation and an autocorrelated deterministic component of utility using residential location choice data for only four zones in the city of Sendai in Japan. Garrido and Mahmassani (2000) recommend using a multinomial probit (MNP) model with spatially and temporally correlated error structure to model (analyze and forecast) the distribution of freight flows over space and time at the operational or tactical levels of planning. They use Monte-Carlo simulation with a choice set of 31-41 location alternatives.

It is worth mentioning here the 2010 paper by Smirnov, which develops a new spatial random utility framework where individual decision-makers are spatially dependent in their preferences. In addition, pseudo maximum likelihood estimator is consistent and computationally feasible for large datasets. The author discusses recent developments of spatial discrete choice models and classifies them by the type of spatial effect and the way it is incorporated in the model.

The literature is scarce on previous attempts to incorporate spatial effects in an establishment's or firm's (re-)location decision process. Vichiensan et al. (2005) confirm that discrete choice models are applied mainly in the field of transport choice analysis and have been continuously developed for years. However, less attention has been put into the research and development of location choice models. Location choice models differ substantially from transport choice models given that they use geo-referenced data, for which choice sets have an explicit spatial component. For this reason, it is important to understand and represent spatial

correlation in location choice models. Nguyen et al. (2012) offer a three-stage firm relocation process: a firm first takes a decision on whether to stay in its actual location or to move, it then chooses the region and, finally, chooses the zone in which to relocate. The authors incorporate spatial interactions among firms in the deterministic part and spatial correlation between zones in the error term (the generalized autoregressive term is used to explain the spatial correlation between the zones). Nevertheless, certain aspects of their research paper can be further improved: for instance, the accessibility of the region used in the analysis is estimated, following Allen et al. (1993), as a linear function whereas a non-linear function could have been proposed instead. Klier and McMillen (2008) use GMM spatial logit model that can be applied to the large samples to account for the clustering of auto supplier establishments in the US.

Thus, in spite of recognizing the importance of incorporating spatial interactions in location choice models (and establishment location choice models in particular), research is insufficient on this topic due to lack of data at the micro level, to numerous alternatives in location choice processes, and to difficulties in defining and representing the spatial effects and in measuring spatial dimensions (Sener et al., 2011). Inspired by spatial econometrics techniques (e.g., Jayet, 2001 and Anselin, 1988, 2003) and drawing from a rich database, we try to find some solutions to the above-stated problems and present our ideas in this paper. We wish to combine the methods used to produce location choice models with spatial econometric techniques by examining the role of space in in these models. This is a new and challenging field of research.

3. Study area

In this paper, we concentrate our research on the Paris region, also known as Ile-de-France. Ile-de-France is a vibrant and innovative region with over 5,6 million jobs, 37% of French executives, and 40% of national workforce in research and development⁴. Yet, the Paris region's economy is also spatially unbalanced (Combes et al., 2011). Ile-de-France is divided into eight *départements* - French administrative units consisting of 1300 municipalities that cover the city of Paris and its suburbs⁵. While the Paris region represents only 2.2% of the surface of France, over 19% of the country's population reside in this area (11.7 million)⁶. The GDP of the region amounts to 29% of total French GDP (IAU IdF, 2014). Large differences in population and employment densities are to be found between Paris and its outer periphery. As illustrated during the 2011 OECD Meeting, many issues continue to stand in the way of improving the situation in the suburbs (de Palma, 2011), such as poor access to public services and infrastructure.

The Grand Paris Project aims to link major territorial development contracts raising the attractiveness of the entire Paris metropolitan region. Estimates are that it will attract 1.5 million people, create or relocate 1 million jobs by 2030 and cost around 32.4 billion euros. Urban planners are likely to connect economically

⁴See <http://www.iau-idf.fr/lile-de-france/un-portrait-par-les-chiffres/lile-de-france-capitale-de-rang-mondial.html>

⁵The Paris region consists of the Paris City, the inner ring with *départements* 92, 93, 94, and the outer ring with *départements* 77, 78, 91, and 95.

⁶Data for the year 2009. Source: <http://www.iau-idf.fr/lile-de-france/un-portrait-par-les-chiffres/population.html>.

promising areas; yet, they could also use transport infrastructure to link the physically deteriorated, poorly planned areas in a bid to stimulate their development as well.

4. Data

Many different data sources were compiled for the present study, drawn primarily from the Census survey of establishments for the years 2003-2011⁷. Changes in nomenclature and definitions, as well as availability of data, meant that the year 2007 was used to run the models and generate the descriptive statistics. Data are pooled across activity sectors and additional data are drawn from other sources to build the final sample we used for estimation.

Local population and workforce are described by various characteristics, such as socio-professional class and education level. These data were gathered by the General Census and provided by the French National Institute of Statistics and Economic Studies (INSEE) for the years 2006-2009. French National Census of Population (RGP) provides us with information on the total number of trips made between work and home for the years 2006-2009. The Public Finances General Directorate provided information on income imputation and tax levels for 1998-2008. We also collected data on the rate of business and residence taxes for the years 2002-2010. Data on business taxes were only available for about half of the analyzed municipalities and the missing 563 out of 1300 values could not be implemented. For this reason, the information on business taxes was not used to run the models. On the other hand, data on residence taxes paid per municipality were available for all the municipalities.

The Callon database enabled us to gather detailed information on the average price and rent levels per square meter for new, renovated, and pre-existing non-renovated offices, for first class shops (the most exclusive shops), for second and third class shops (the least exclusive shops), as well as for individual industrial units and warehouses. Data on real estate prices were collected for the years 2000-2009. The hedonic pricing method was used to impute the missing values for some municipalities.

From the Sitadel2 database, managed by the Ministry of Ecology, Sustainable Development and Energy, we drew information on housing construction (net floor area and number of units for housing and professional offices) for the years 2002-2011. The numerical land use database (Corine Land Cover), also managed by this Ministry, was used to gather information on urban and agricultural zones for 2000 and 2006. The MOS database of the Urbanism and Development Institute enabled us to obtain information on the portion of the municipality areas dedicated to shops, services, industrial activities, universities and schools, hospitals and clinics, and various other amenities, for example, parkings and parks. The same database gave us the percentage of municipalities' surface covered by residential areas or vacant land. In addition, the BPE equipment database of DSDS/INSEE provided us with information on the location of individual services,

⁷The Census survey was carried out yearly until the year 2011 by the French National Institute of Statistics and and Economic Studies.

education, health, social actions, transport services, sport and leisure zones, etc. Finally, information on the public transport accessibility level and the proximity to highways for 2005-2009 was obtained through the Regional and Interdepartmental Directorate of Public Works and Planning.

Table 1: Description of potential explanatory variables

Variable and its expected sign	Description
Establishments from respective sector (+) ^a	Number of pre-existing establishments from the analyzed sector within a particular municipality divided by the surface of municipality (km2) ^b
Large establishments from all sectors (+,-)	Number of large pre-existing establishments with fifty employees or more divided by the surface of municipality (km2)
Employment density (+)	Size of active employment (x1000) divided by the surface of municipality (km2)
White-collar employees (+)	Number of white-collar workers divided by the size of labor force
Blue-collar employees (+)	Number of blue-collar workers divided by the size of labor force
Trips home-work (nl)	Number of trips between home and work if municipality is both a place of residence and a workplace to the total number of trips home-work
Population density (+,-)	Size of population (x1000) divided by the size of residential area (km2)
Income per person (+)	Average income level per capita (euros)
Offices (+)	Fraction of a municipality's surface dedicated to offices
Shops (+)	Fraction of a municipality's surface dedicated to shops
Parkings (nl)	Fraction of a municipality's surface dedicated to parkings
Universities and schools (+)	Fraction of a municipality's surface dedicated to universities and schools
Hospitals and clinics (nl)	Fraction of a municipality's surface dedicated to hospitals and clinics
Vacant land (+)	Fraction of a municipality's vacant land available for new investments
Distance to highway (-)	Distance to the nearest highway (km)
Public transport (+)	Number of subway, train stations, and bus stops in a municipality
Residence tax (-)	Average level of residence taxes
Price of offices (-)	Average price level of offices per square meter (euros)
Price of shops (-)	Average price level of shops per square meter (euros)

^a(+) and (-) means that the associated coefficient is expected to be positively or negatively statistically significant, respectively. (nl) means that no literature treats this problem or that no literature was reviewed on this issue.

^bThe data on stock of establishments are given for the 1st of January 2007. The range for the independent variables is 2005-2009.

Based on the reviewed literature, the paper of Arauzo-Carod et al. (2010) and the DREIF report of de Palma et al. (2008) in particular, the available data, and our prior knowledge, we select the variables that can be used in the models to explain the location choice of newly created establishments. We run the models with variables that are expected to represent the structures of the population and employment area, the measure of accessibility to specific populations and employment characterized by certain relevant skills. We test whether the location choices of newly created establishments are influenced by the proximity of retail, services, universities and schools, and other amenities. In addition, the proportion of a municipality's vacant land that can be available for new investments may play a role as well as the proximity to residential areas. The significance of price of shops and offices on the location choice of establishments is tested. We also

analyze, whether an easy access to a particular municipality by either public or private transport influences the location choices, and for which sectors these accessibility measures are most relevant.

When modeling the location choice of establishments, care should be taken to select appropriate set of variables depending on the activity sector being analyzed. The expected signs of the estimated parameters may differ accordingly. As seen in the review by Arauzo-Carod et al. (2010) and the literature review presented in subsection 2.1, the focus has been, up to this point, on the manufacturing/industrial and high-tech or R&D sectors. We here consider several other sectors, such as construction, hotels and restaurants, and real estate activities. Thus, the existing literature can be treated as a guide to motivate factors that may be studied in the econometric models. One should be cautious, however, interpreting the models' results, as, for example, the signs of the obtained estimates depend on the sector analyzed.

The data used to estimate the models are gathered at the municipality level. We also used these raw data to compute additional variables for our application. When constructing the database, we paid special attention to the problem of definition changes, differences within the available periods and appropriate aggregation, interpolation, and extrapolation techniques. Selected variables that have been tested in the models along with their estimates expected signs are summarized in Table 1.

5. Descriptive statistics

To carry out the present study, we accessed the Census data of establishments at the aggregate level for the years 2003-2011. Due to changes in nomenclature, activities in 2003, 2005, 2007, 2009, 2010, and 2011 are based on the NAF nomenclature (the French classification of activities) while activities in 2006 and 2008 are based on the NES nomenclature (the summary economic classification). There is no one-to-one correspondence that would allow us for the comparison of all sectors or sub-sectors between these two classifications. Definitions of some sectors used in NES and NAF nomenclatures change across time. In addition, on August 4th 2008, the *Loi de Modernisation de l'Economie* (LME)⁸ was introduced in an attempt to stimulate growth and employment creation. It gave small and medium enterprises the best opportunities to grow by reducing their payments, boosting competition (e.g., simplifying the installation of supermarkets) and enhancing the attractiveness of the territory. On January 1st 2009, the definition of the auto-entrepreneur status was also modified⁹. Taking into account all these modifications and data accessibility, we built the models for the year 2007. Thus, 2007 is the year in which the dependent variable is measured.

Data are pooled across eleven main activity sectors: 1) industry (Industr), 2) construction (Constr),

⁸More information on LME can be found on: <http://www.toute-la-franchise.com/vie-de-la-franchise-A3942-commerces-de-centre-ville-la-muta.html>.

⁹It applies to the natural persons who set up or already possess a sole proprietorship, for the purpose of exercising a commercial or artisanal activity or one of the professions (with the exception of certain activities), as a main or complementary activity, and whose sole proprietorship fulfills the conditions of the micro-enterprise fiscal category, and who opt for VAT exemption. Source: <http://www.insee.fr/en/methodes/default.asp?page=nomenclatures/liste-nomenclatures.htm>.

3) commerce (Commer), 4) transport (Transp), 5) financial and insurance activities (Finan), 6) real estate activities (RealEst), 7) hotels and restaurants (HoteRes), 8) information and communication (InfoCom), 9) special, scientific and technical activities (SpecSci), 10) education (Educ), 11) health and social actions (HeaSoc), and the additional category "others"¹⁰.

5.1. Establishments' structure across activity sectors

In 2007, the largest number of newly created establishments (over 22%) belonged to the commerce sector. The percentage of pre-existing commercial establishments among all establishments (also 22%) had been decreasing across time, when compared to 2003 and 2005. The second biggest sector was the special, scientific and technical activities sector, with 17.4% of newly created units and 16.2% of pre-existing establishments. The percentage of establishments in this sector increased across time. The third biggest sector was construction, with 13.8% of newly created units and a much lower 9.2% of pre-existing establishments. We also observed the smallest sectors: education (2% for newly created establishments and 1.6% for pre-existing establishments) and transport (3.5% for newly created establishments and 3.9% for pre-existing establishments).

5.2. Ranking of municipalities

Next, we analyze the numbers of newly created establishments across all 1300 municipalities. We compute the market share of each municipality as the number of newly created establishments in a given municipality from a given sector s over the total number of newly created establishments from this sector s . This allows us to build the rankings of municipalities based on their share in the market. This method allows us to see which municipalities and *départements* are the most frequently chosen for a new location. From the rankings of municipalities, we select the municipality (one from each activity sector) which attracts the highest number of newly created establishments. We also determine the number of new units that locate in the top 10, 20, 50, and 100 municipalities. This allows us to offer some insights on the concentration of activities in the Paris region (see Table 2). For instance, the most frequently chosen municipality, namely the 8th district of Paris, becomes home to over 14% of newly created units in the real estate sector and it also accounts for 13.4% of all newly created financial establishments. Approximately 9.3% of all new establishments in the special, scientific and technical activities sector are also located in just one of the 1300 potentially available municipalities (the 8th district of Paris). On the other hand, we observe relatively low fractions of transport and commerce newly created establishments in the most frequently chosen municipality: only 2.3% and 2.7%, respectively. As a means of contrast, the average percentage of establishments that locate in the top

¹⁰"Others" includes: administration services; public administration; arts, spectacles and recreation activities; other service activities: activities of membership organizations, reparation of computers and personal and household goods, and other personal services; activities of households as employers of domestic staff and activities of households as producers of goods and services for their own use; and extra-territorial activities.

Table 2: Market share of top 1, 10, 20, 50, 100 municipalities in 2007

#Municipalities ^a	Industr ^b %	Constr %	Commer %	Transp %	Finan %	RealEst %
1	4,4 % (146)	4,2% (505)	2,7% (531)	2,3% (70)	13,4% (595)	14,4% (672)
10	25,6% (842)	21,8% (2644)	22,2% (4360)	17,6% (541)	39,9% (1772)	40,4% (1892)
20	37,9% (1249)	33,1% (4004)	34,7% (6815)	29,3% (900)	52,3% (2327)	52,9% (2478)
50	56,1% (1849)	51,7% (6266)	53,6% (10541)	52,1% (1600)	66,6% (2963)	68,2% (3195)
100	69,5% (2291)	67,8% (8211)	68,8% (13533)	70,8% (2175)	77,7% (3455)	78,0% (3654)
All municipalities	3296	12115	19658	3072	4446	4683
Sector size	3,8%	13,8%	22,3%	3,5%	5,1%	5,3%
#Municipalities	HoteRes %	InfoCom %	SpecSci %	Educ %	HeaSoc %	All %
1	3,6% (125)	5,4% (349)	9,3% (1427)	3,8% (66)	3,7% (149)	5,1% (4525)
10	25,4% (896)	31,7% (2048)	36,6% (5587)	26,9% (466)	24,7% (1008)	26,1% (23004)
20	38,6% (1359)	47,6% (3076)	52,4% (7984)	40,5% (702)	36,4% (1484)	38,6% (33996)
50	58,1% (2046)	67,4% (4349)	69,2% (10582)	58,3% (1011)	53,6% (2185)	56,9% (50026)
100	73,2% (2581)	79,3% (5120)	79,2% (12103)	72,2% (1251)	68,8% (2804)	70,2% (61780)
All municipalities	3524	6457	15282	1733	4077	87974
Sector size	4,0%	7,3%	17,4%	1,97%	4,6%	100%

^aNumber of top 1, 10, 20, 50, 100 municipalities which gather the highest number of newly created establishments from a particular activity sector in 2007.

^bMarket share of top municipalities. Number of newly created establishments from a particular activity sector within these top municipalities given in the bracket.

municipality is 5.1%. The first 20 municipalities in the real estate, financial, scientific, special and technical activities, and information and communication sectors account for around half of all newly created units.

In addition, the top municipality in the case of the financial activities sector (the 8th district of Paris) gathers 2.6 times more newly created establishments than the second top municipality, the 16th district of Paris. Similarly, the same top municipality (the 8th district of Paris) attracts 2.3 times more newly created real estate units than the "second best" municipality, the 16th district of Paris. A large difference in the number of newly created establishments between the two first municipalities from the ranking lists can also be observed in the construction sector. In this sector, establishments tend to locate 1.7 more often in the 20th district of Paris than in the "second best" municipality, the 10th district of Paris. The top municipalities (based on market share) are the ones that belong to Paris and to *départements* 92 and 93. However, some differences are observed across sectors. We summarize the results on the maps in Figure 1.

5.3. Size of newly created establishments versus size of pre-existing units

We concentrate now on the size of newly created units compared to the size of pre-existing establishments (see Table 3). Around 60% of all pre-establishments are units with zero employees. An establishment with zero employees is an operating establishment with no employees other than the owner. Establishments with 0-9 employees (so called micro establishments) account for 93% of all establishments. In comparison, these figures stand respectively at 82% and 97% for newly created establishments. Next, 7% of pre-existing units

are registered as establishments with 10 or more employees, compared with only 2.8% of newly created units from the same size category. Interestingly, 1.4% of existing units have at least 50 employees, whereas it is 7 times less frequent for newly created establishment to be classified to this category. Thus, a newly created establishment will most probably be one with zero employees and only very rarely it will have more than 10 employees. Stable units grow in size and hire more workers.

Table 3: Structure of newly created and pre-existing establishments by size category

New	#Estab 0 ^a	%	#Estab 1-9 ^b	%	#Estab 10+ ^c	%	#Estab 50+ ^d	%	Total
Industr	2442	74,1 %	628	19,1 %	226	6,9 %	23	0,70 %	3296
Constr	8554	70,6 %	2505	20,7 %	1056	8,7 %	3	0,02 %	12115
Commer	15472	78,7 %	3810	19,4 %	376	1,9 %	29	0,15 %	19658
Transp	2496	81,3 %	491	16,0 %	85	2,8 %	13	0,42 %	3072
Finan	3855	86,7 %	503	11,3 %	88	2,0 %	26	0,58 %	4446
RealEst	4049	86,5 %	599	12,8 %	35	0,7 %	7	0,15 %	4683
HoteRes	2013	57,1 %	1384	39,3 %	127	3,6 %	8	0,23 %	3524
InfoCom	5812	90,0 %	559	8,7 %	86	1,3 %	11	0,17 %	6457
SpecSci	14099	92,3 %	1040	6,8 %	143	0,9 %	23	0,15 %	15282
Educ	1589	91,7 %	133	7,7 %	11	0,6 %	1	0,06 %	1733
HeaSoc	3916	96,1 %	132	3,2 %	29	0,7 %	6	0,15 %	4077
Other	8028	83,4 %	1380	14,3 %	223	2,3 %	31	0,32 %	9631
Total	72325	82,2 %	13164	15,0 %	2485	2,8 %	181	0,21 %	87974
Stock	#Estab 0	%	#Estab 1-9	%	#Estab 10+	%	#Estab 50+	%	Total
Industr	20872	46,6 %	16800	37,5 %	7072	15,8 %	1694	3,79 %	44744
Constr	36390	52,1 %	28622	40,9 %	4896	7,0 %	589	0,84 %	69908
Comm	92715	55,1 %	63047	37,5 %	12473	7,4 %	1930	1,15 %	168235
Transp	19624	65,4 %	6213	20,7 %	4183	13,9 %	1071	3,57 %	30020
Finan	23853	61,4 %	11404	29,4 %	3590	9,2 %	892	2,30 %	38847
RealEst	31325	72,8 %	10567	24,6 %	1139	2,6 %	195	0,45 %	43031
HoteRes	16837	35,8 %	24621	52,3 %	5577	11,9 %	521	1,11 %	47035
InfoCom	30536	66,4 %	11121	24,2 %	4346	9,4 %	1139	2,48 %	46003
SpecSci	88778	71,9 %	28419	23,0 %	6268	5,1 %	1189	0,96 %	123465
Educ	9079	73,2 %	2583	20,8 %	749	6,0 %	94	0,76 %	12411
HeaSoc	50789	84,1 %	8365	13,8 %	1266	2,1 %	371	0,61 %	60420
Other	47052	59,6 %	26490	33,5 %	5470	6,9 %	1202	1,52 %	79012
Total	467850	61,3 %	238252	31,2 %	57029	7,5 %	10887	1,43 %	763131

^a#Estab 0: Number of establishments with 0 employees.

^b#Estab 1-9: Number of establishments with 1-9 employees.

^c#Estab 10+: Number of establishments with 10 employees or more.

^d#Estab 50+: Number of establishments with 50 employees or more.

Large discrepancies in the typical size of newly created units can be observed across sectors. 96% of all newly created establishments in the health and social actions sector and 92% in the education and special, scientific and technical activities sectors have zero employees. These figures can be compared to the 57% of

newly created hotels and restaurants with zero employees. Relatively high figures of larger establishments with more than 10 employees are registered in the construction and industry sectors (8.7% and 6.9%, respectively) in comparison to rather low levels for other sectors. New creation of very large units (with 50 employees or more) is in general not common (from nearly zero percent for construction to 0.7% for industry) and is most likely to occur within the industry, financial activities, and transport sectors.

6. Parametric models

6.1. Count data models

In an economic context, the dependent variable is often an integer or count with the distribution that places probability mass at nonnegative integer values only (unlike a regression model). The Poisson regression is a very popular model and usually forms the starting point for count data analysis. Such a model captures the discrete and nonnegative nature of the data; it does, however, assume equidispersion (equality of the conditional variance and the conditional mean). This assumption is found to be too restrictive given that the conditional variance often exceeds the conditional mean. Large overdispersion leads to deflated standard errors and inflated t-statistics in the maximum likelihood output. Thus, a statistical test of the overdispersion should be performed after running the Poisson model. Overdispersion can be modeled either by introducing an additional parameter as in the negative binomial (NB) model (Consul and Jain, 1973), by allowing for an extra proportion of zeros using zero inflated models or by combining zero inflated models with overdispersion distribution.

In addition, when the observed data display a higher fraction of zeros than would be typically explained by the standard count data models, two types of models can be suggested: the hurdle model (Mullahy, 1986) or the zero-inflated model (Lambert, 1992). Where the Poisson model is used, the density function predicts the probability of zeros to be less than that is actually observed.

The hurdle model, also called the two-part model, reflects a two-part decision making process. It relaxes the assumption that the zero observations and the positive observations come from the same data generating process. Maximum likelihood estimation of the hurdle model involves separate maximization of the two terms in the likelihood, one related to the zeros and the other to the positive values. Typically, the distribution of the second part has a zero-truncated form of a standard discrete distribution such as Poisson or NB distribution, although other distributions defined directly on the positive integers (e.g., the logarithmic) can be also considered. The two-stage decision-making process is reflected through the hurdle model interpretation. A zero-inflated model assumes that there are two sources of zero counts (not just one as in the hurdle model). For an extensive description of count data models and their comparison and applications, see e.g., Winkelmann (2003) or Cameron and Trivedi (1986).

In the current paper, we run several models, namely: Poisson, zero-inflated Poisson (ZIP), zero-inflated tau Poisson (ZIP tau), negative binomial (NB), zero-inflated negative binomial (ZINB), and hurdle models.

All these models are run separately for selected activity sectors¹¹: 1) not incorporating spatial effects and 2) using distance matrix. The results are then compared and the best performing hurdle model with incorporated spatial spillovers is described in the next section.

We agree that we may have used other modeling techniques by transforming the dependent variable into a rate instead of a count, using, for instance, an aggregate logit model. We, thus, would have had to control for the excess of zeros problem (i.e. rate equals to 0) by means of an additional sample selection equation (like a Tobit model applied to rates). We have no prior about which approach fits best the data. We leave this aside for future research. Yet, we have chosen to use an approach where it is easier to control for the excess zeros, as presented in the next section.

6.2. Hurdle-Poisson model

We statistically describe the location of newly created establishments that belong to a given economic sector s in zone l at the aggregate level. Let $y_{l,s}$ models the count of newly created establishments from sector s that locate at l . By construction, $y_{l,s}$ takes nonnegative integer values. The rationale behind the choice of a hurdle-Poisson model is that observing a zero outcome means that it is not possible to capture additional net profit at a given location l for sector s provided set establishing conditions at the beginning of the year: there is no reason to locate where there is no profit to make. New establishments will locate in a zone where there remains profit to be captured. We agree that this is a debatable issue, as we do not account for behavioral dynamics (forward-looking behaviors) of establishments. However, we prefer to stick to the following: no current additional profit, no new establishment. We also understand that already installed establishments may increase in size and/or in production to capture any additional profit.

We describe observations of these counts by independent variables $\mathbf{x}_{l,s}$. As our empirical approach is a two-part model, these variables may be used in either of the parts. They may also differ along with s . Our approach is parametric: let β_s (hurdle part) and α_s (conditional count part) be the weights of the independent variables.

The contribution of one observation to the sample likelihood function is written as (Cameron and Trivedi, 1986):

$$\begin{aligned} \ell(\alpha_s, \beta_s | y_{l,s}; \mathbf{x}_{l,s}) = & \\ \Pr(y_{l,s} = 0 | \mathbf{x}_{l,s}; \beta_s)^{\mathbb{I}(y_{l,s}=0)} \times & \\ ((1 - \Pr(y_{l,s} = 0 | \mathbf{x}_{l,s}; \beta_s)) g(y_{l,s} | y_{l,s} > 0; \mathbf{x}_{l,s}; \alpha_s))^{1 - \mathbb{I}(y_{l,s}=0)}. & \end{aligned} \quad (1)$$

We obtain an analytical formulation of the model by assuming that g is a truncated-at-zero Poisson distribution:

$$g(y_{l,s} | y_{l,s} > 0; \mathbf{x}_{l,s}; \alpha_s) = \frac{\mu(\mathbf{x}_{l,s}; \alpha_s)^{y_{l,s}} \exp(-\mu(\mathbf{x}_{l,s}; \alpha_s))}{y_{l,s}! (1 - \exp(-\mu(\mathbf{x}_{l,s}; \alpha_s)))}, \quad (2)$$

¹¹These seven selected sectors are: industry, construction, commerce, financial activities, real estate activities, hotels and restaurants, and special, scientific and technical activities.

where the rate μ is defined as:

$$\mu(\mathbf{x}_{l,s}; \alpha_s) = \exp(\mathbf{x}'_{l,s} \alpha_s). \quad (3)$$

In addition, a logistic distribution is used to characterize the hurdle:

$$\Pr(y_{l,s} = 0 | \mathbf{x}_{l,s}; \beta_s) = \frac{\exp\left(-\frac{\Pi(\mathbf{x}_{l,s}; \beta_s)}{\sigma}\right)}{1 + \exp\left(-\frac{\Pi(\mathbf{x}_{l,s}; \beta_s)}{\sigma}\right)}, \quad (4)$$

where:

$$\Pi(\mathbf{x}_{l,s}; \beta_s) = \mathbf{x}'_{l,s} \beta_s. \quad (5)$$

Equation 5 models the deterministic part of a latent profit function. The probability that there is no establishment of type s which locates at l increases when the potential (latent) profit decreases.

It is straightforward to derive that the expectation of the count variable is given by:

$$\mathbb{E}(y_{l,s} | \mathbf{x}_{l,s}; \alpha_s, \beta_s) = \Phi(\mathbf{x}_{l,s}; \alpha_s, \beta_s) \mu(\mathbf{x}_{l,s}; \alpha_s) \quad (6)$$

and that the variance is equal to:

$$\mathbb{V}(y_{l,s} | \mathbf{x}_{l,s}; \alpha_s, \beta_s) = \Phi(\mathbf{x}_{l,s}; \alpha_s, \beta_s) \mu(\mathbf{x}_{l,s}; \alpha_s) (1 + (1 - \Phi(\mathbf{x}_{l,s}; \alpha_s, \beta_s)) \mu(\mathbf{x}_{l,s}; \alpha_s)), \quad (7)$$

where:

$$\Phi(\mathbf{x}_{l,s}; \alpha_s, \beta_s) = \frac{1 - \Pr(y_{l,s} = 0 | \mathbf{x}_{l,s}; \beta_s)}{1 - \exp(-\mu(\mathbf{x}_{l,s}; \alpha_s))}. \quad (8)$$

Overdispersion occurs when $\Phi(\mathbf{x}_{l,s}; \alpha_s, \beta_s) < 1$ and underdispersion when $\Phi(\mathbf{x}_{l,s}; \alpha_s, \beta_s) > 1$.

Another interesting feature of the model is that the total effect of an independent variable on the count process passes through two channels. Looking at the elasticity of the expectation of the count variable $y_{l,s}$ to any related variable $x_{l,s}$, we find that it is equal to the sum of the elasticity of the participation probability to this variable and the elasticity of the conditional expected count to this variable:

$$\epsilon_{\mathbb{E}(y_{l,s} | \mathbf{x}_{l,s}; \alpha_s, \beta_s) / x_{l,s}} = \epsilon_{1 - \Pr(y_{l,s} = 0 | \mathbf{x}_{l,s}; \beta_s) / x_{l,s}} + \epsilon_{\mathbb{E}(y_{l,s} | y_{l,s} > 0, \mathbf{x}_{l,s}; \alpha_s) / x_{l,s}}. \quad (9)$$

Given a sample of independently and identically distributed observations for l locations and s sectors, the log-likelihood function is written as:

$$\ell(\alpha_s, \beta_s | \mathbf{y}; \mathbf{x}) = \sum_{l=1}^L \sum_{s=1}^S \ell(\alpha_s, \beta_s | y_{l,s}; \mathbf{x}_{l,s}). \quad (10)$$

Note that we here assume that there are no unobserved correlations between different sectors s . We also do not account for spatial errors and spatial lags. There is a need to fill this gap in the further studies. The paper of Lambert et al. (2010) and their spatial lag model of counts can give a hint.

6.3. Spatial spillovers

We discuss here the structure of the matrix of observed explanatory variables. $\mathbf{x}_{l,s}$ is actually a generic notation used for the sake of conciseness in notations. Due to data availability and our prior knowledge

regarding the various expected effects, this matrix has a specific structure: it contains implicitly the number of pre-existing establishments from a respective sector s located at l , the number of large pre-existing establishments from all sectors located at l , and other explanatory variables directly concerning either location l or sector s . It is also possible to make this structure more specific by accounting for spatial interdependence: the stocks of establishments from some sectors located in neighboring cities and locational-specific attributes of neighbor cities.

Without going too deeply into the discussion, we will simply state that we make use of the distance matrix to characterize spatial patterns¹². Currently, there are two basic categories that define neighbors: contiguity (shared borders) and distance. Contiguity-based weights matrices include rook and queen matrices. A rook matrix defines a neighbor as an area with a shared border while a queen matrix defines a neighbor as an area with a shared border and a shared vertex (point) (e.g., on a grid); in addition to the four cells included under a rook matrix, the four cells sharing a corner with the central location are also counted as neighbors. Distance-based weights matrices include distance bands and k nearest neighbors¹³.

Spatial spillovers are simply modeled as:

$$\mathbf{x}_{l,s} = \ln \left(\sum_{j=1}^L e^{-\mu d_{l,j}} \mathbf{z}_{j,s} \right), \quad (11)$$

where $\mathbf{z}_{j,s}$ is an attribute of the municipality that applies to sector s (the same attribute may apply to several sectors) or is the number of pre-existing establishments from this sector. $\mu = 1$ is fixed to 1 for our application. There is further work to carry out about values that μ may take. As it models the range of spatial spillover effects, it may at least differ along with sectors s . This work is left aside for the moment, yet, we refer the reader to Buczkowska et al. (2013) to have an idea about potential consequences $d_{l,j}$ is the distance between the centroids of municipalities l and j .

No period-specific pattern is modeled. Nevertheless, we implicitly parameterize a dynamic structure in our cross-sectional data by assuming that the location of newly created establishments during our period of observation is related to the stock of pre-existing establishments at the beginning of said period. The interdependence between activity sectors can be included by considering that newly created establishments within one activity sector are related to the stock of pre-existing establishments from both said and other activity sectors. As our geographical structure is based on zones, we need to consider that there may exist some spatial patterns between them.

¹²Rook or queen contiguity matrices could have been proposed as alternatives.

¹³Source: <https://geodacenter.asu.edu/node/390>.

7. Results

7.1. Count data model comparison

We run Poisson, zero-inflated Poisson, zero-inflated tau Poisson, negative binomial, zero-inflated negative binomial, and hurdle models. Each model is run separately for each of the selected activity sectors. Each model is specified along two dimensions: 1) without accommodating spatial effects and 2) using the distance matrix. We run the models with variables that are expected to represent the municipality characteristics and the numbers of pre-existing establishments.

Some of the models we run are nested (e.g., Poisson and NB, Poisson and hurdle) while others are non-nested (e.g., ZIP and Poisson, ZIP and NB, ZIP and hurdle). Two models are nested if one can be reduced to the other by imposing a set of linear restrictions on the parameter vector. Conversely, two models are non-nested (either partially or strictly) when this condition is not met (Clarke, 2003). In the statistics literature, two methods are generally used to compare non-nested models: the Vuong test proposed by Greene (1994) and the Clarke test (2003). Non-nested models can still be compared using information criteria, such as AIC (Akaike Information Criterion) and BIC (Schwartz' Bayesian Information Criterion) provided in most procedures using maximum likelihood estimation. Models with smaller values of these criteria are considered better models. However, no statistical test comparing criterion values is available.

The main conclusion we draw from our results is that incorporating spatial spillovers between location choice alternatives (municipalities) in the form of a distance matrix and for each type of model and for all the activity sectors¹⁴, improves the estimates. Looking at the models for which the distance matrix was used and postulating that the justification of best model is solely based on statistical tests, it would be tempting to conclude that NB outperforms its counterparts for lower AIC and BIC. The NB model outperforms the Poisson model for each activity sector, as it is characterized by a lower level of the absolute value of log-likelihood (for nested models, the closer to zero the level of log-likelihood, the better the model). Yet, hurdle and ZIP regressions offer greater flexibility in modeling zero outcomes than do the other count data models that we run. The major difference between hurdle and ZIP models is that the logit component of the hurdle model estimates the probability of a zero count, whereas the logit component of the ZIP model describes the probability of a zero count from two groups of either always zero and not always zero. Thus, the interpretation of the hurdle model is more straightforward and more intuitive than that of the ZIP model. The first part of

¹⁴We test six types of count data models: Poisson, ZIP, ZIP tau, NB, ZINB and hurdle models for seven selected activity sectors with variables that represent the municipality's characteristics, the density of pre-existing establishments from the respective sector, and the density of large pre-existing establishments. Each model is then run with and without incorporating spatial spillover effects between the municipalities. In this way, we are able to obtain results from 84 models. In a mere four out of all analyzed cases (that is to say, cases for which running the model led to function convergence) did the results indicate that there was no need to include spatial spillovers in the models. These four exceptions are: the hurdle model for the industrial establishments (very similar levels of log-likelihood, AIC, and BIC criteria were obtained when running the models with and without the distance matrix); the ZIP tau model for the industry, construction, and real estate activities sectors.

the estimates in the hurdle model explains why no new establishments are created in a given municipality, whereas the second part of the estimates describes the influence of various determinants on the positive creation of units within the municipality. Having no creation of establishments in a municipality happens when profit is exhausted (for instance by other pre-existing establishments within the municipality) or when there is no space for new units.

We observe that the ZIP model outperforms the ZIP tau model in all but one case¹⁵. The hurdle model outperforms the ZIP tau model in all of the analyzed cases, with no exceptions. A hurdle model usually gives better results than a ZIP model¹⁶. For these reasons, ZIP and ZIP tau models are not treated as rivals for the hurdle model. ZINB could have been suggested as an alternative to the hurdle model since each sector is characterized by very low values of the AIC and BIC criteria (lower than in the case of hurdle and ZIP models). However, when running the ZINB model, whether using the quasi-Newton optimization method or not in the case of many sectors, convergence could not be reached under the requirement of estimating all parameters. Taking into account all of these considerations, we therefore chose favor the hurdle model.

7.2. Hurdle model: effect of various municipality characteristics on location choice

In the highly heterogeneous Paris region, some municipalities find themselves home to a large number of new establishments whereas others struggle to be chosen by any establishment. The first part of the hurdle estimates shows the reasons for no newly created establishment in the municipality. Depending on the analyzed sector, the percentage of municipalities left with no new creation ranges from 34% up to 69%¹⁷. In the subsection 7.2.1, we present an interpretation of the results from the first part of the hurdle models. In the following subsection (7.2.2), the reader can find an interpretation of the results from the second part of the hurdle models with a positive number of establishment creations occurring in the municipality. For such location choice models and for all analyzed sectors, we included a variable to represent "the number of pre-existing establishment", which is in addition to variables representing the characteristics of, and access to, the municipality.

Regarding the weight matrix, the spatial modeling accounts for the density of establishments from the respective sector in neighboring areas, the employment density of white- and blue-collar workers, and also the existence of large employers from all sectors with fifty workers or more. These large establishments might attract complementary upstream and downstream establishments (sometimes even within the same firm), leading to positive autocorrelation, or which might affect the labor market in such a way that it

¹⁵This exception is the model run without the distance matrix for the industry sector.

¹⁶In the case of the ZIP model when no matrix is used, for the real estate and special, scientific and technical activities sectors, not all parameters were estimated.

¹⁷The number of municipalities left with zero newly created establishments in the industry sector equals to 734, in construction 439, commerce 440, transport 837, financial activities 799, real estate activities 738, hotels and restaurants 792, information and communication 771, special, scientific and technical activities 569, education 890, health and social activities 794 out of 1300 possible municipalities.

becomes costlier to hire employees and locate nearby, leading to negative autocorrelation. For this reason, it is necessary to capture the possible influence of large establishments on the location choice of newly created units.

7.2.1. No new creation of establishments in municipality

Large establishments with fifty employees or more seem not to add much in the explanation of why there are no new establishment openings in the area, beside that they have a weak negative effect on the real estate sector and a positive effect on the financial sector. A municipality is left with zero new openings of industrial establishments when there are no, or a few other establishments from the industry sector and/or when the average level of real estate prices is too high. Relatively few pre-existing construction establishments, low population density, high office prices, and/or difficulties in accessing public transport do not seem to encourage the creation of new establishments in the construction sector. There is a propensity for no new commercial establishments to be created in municipalities that lack other commercial units and/or when commercial rents are particularly high, both within the area and its surroundings. No new creation of hotels and restaurants can be observed in the areas characterized by steep price level, not enough vacant land where a new investment could be made, and/or few pre-existing units from this sector. No new financial establishments will be created when the number of pre-existing establishments within that same sector is too low. Steep office prices and limited access to white-collar workers, all beyond a certain threshold, may also lead to an absence of new openings for financial units. Difficulties in accessing the potential location by public transport may leave the municipality without any new openings of financial establishments either. There are also various reasons behind a municipality being left without any new real estate sector establishments. It is most likely to happen when there are too few, or indeed no pre-existing real estate units. It can also occur in case of high residential rate in the area, and/or when access to shops and services is hampered. Finally, a lack of other pre-existing establishments in special, scientific and technical activities are the most significant variables that cause zero new unit creation in this sector. In addition, difficulties for white-collar workers to access the premises, excessive office prices, and difficult access by public transport appear not to help to create a positive number of establishments in the special, scientific and technical activities sector.

7.2.2. Positive number of newly created establishments in municipality

The models presented in this subsection, as well as the models explained in subsection 7.2.1, show the importance of interactions between establishments from a same activity sector. The number of new locations will be highly dependent on the quantity of competing establishments in the municipality and in its surroundings. Pre-existing establishments will encourage new units to locate nearby. This effect is found to be strongest in the case of construction, commerce, and special, scientific and technical activities. Overall, large establishments tend to attract other industrial units, yet, have a negative effect on other analyzed sectors.

Drawing from the results from the model that describes the location choice of industrial establishments, we can observe that units tend to choose a location characterized by easy access to blue-collar workers, large

areas of vacant land that can be made available for new investments, and/or a conveniently located highway. Newly created establishments in the construction sector strongly prefer to locate in an area characterized by a high level of population density. In order to create this variable, we computed the ratio of population size to municipality surface zoned as residential area. We opted to consider only the surface zoned as residential area (instead of the entire municipality's surface) as a means to better express the population density measure. These establishments in the construction sector would rather choose areas with vacant land that could be used for new investments. Another important factor that should be taken into account is close proximity to public transport services. The distance to the highway does not seem to play a very significant role when the decision for a new location is taken. In addition, establishments in the construction sector pay close attention to office prices and residence tax rates. Access to the potential location by cars and close proximity to subway, train stations, or bus stops can be perceived as fundamental factors by a commercial establishment when choosing its location. It is important for future commercial establishments to be located in the vicinity of potential customers. Consequently, establishments will try to reach areas characterized by a high number of population who work and live nearby. Relatively steep shop prices will discourage the opening of new commercial units. Close proximity to subway, train stations, or bus stops seems to be very significant also in the location choice model of hotels and restaurants. Being located at a short distance from the highway will also play a positive, albeit weaker, role when deciding on the new location. Relatively easy access to white-collar workers who can profit from the restaurant or hotel services, availability of vacant land, and low real estate prices can also act as incentives in the location choice.

Establishments in the financial sector will look for municipalities easily accessible by cars. Easy public transport access is also relevant, yet slightly less significant. Financial units favor areas occupied by white-collar workers. It is worth mentioning that office prices appear to be weakly significant when financial activities choose a new location. Existence of shops and offices in and around a chosen municipality seems not to accelerate the new openings in the financial sector. The model for newly created establishments in the real estate sector demonstrates that a potential location should be easily accessible by public transport. Public transit accessibility is the most significant explanatory variables in this model. A potential location should be visible and frequently used. Establishments that belong to the real estate sector favor zones with easily accessible shops and services. Being located in zones that offer "a good image" is usually favored when the establishment chooses its new location. It could be argued that close proximity to wealthy clients attracts new real estate units. Low residence tax rates and low prices seem to strongly encourage new openings in the real estate sector.

Finally, ease of access, be it by public or private transport are significant variables in the model for special, scientific and technical activities. Access to the white-collar workforce within the close vicinity will accelerate the number of new openings in this sector. Yet, these establishments tend to avoid overpopulated areas. Office prices does not play a critical role. Neither the pre-existing offices attract the new establishments in the special, scientific and technical activities sector, indicating that they will rather locate close to other

pre-existing units within the same sector. Tables 4, 5 summarize the results from these estimated hurdle models.

8. Conclusions

The following conclusions are based on both the descriptive statistics and the results from the count data models that we have estimated. Questions such as: Where in the Paris region do new establishments decide to locate most often? and What is the level of activity concentration? are now answerable based on rankings of municipalities' market share. We found, that the top municipalities are the ones that belong to Paris and to *départements* 92 and 93. *Départements* from the outer ring seem to be the least attractive for new establishments. About 14% of all newly created establishments in the real estate and financial and insurance sectors chose to locate in the 8th district of Paris. In contrast, there does not seem to be an ideal municipality, nor one that particularly appeals to the transport and commerce sectors so as to be very frequently chosen as a destination. After determining preference for municipality, our next step consisted in analyzing the size of newly created establishments in comparison to the size of pre-existing ones. We observed a strong contrast between the typical size of these two types of establishments across various size classes. The most appealing conclusion rests on the fact that 82% of newly created establishments have no employees beside the owner whatsoever, whereas this is the case for only 61% of pre-existing establishments. In addition, we detect 7.5% of pre-existing establishments having at least 10 employees, in comparison to the 2.8% of newly created establishment of the same size. Just 0.2% of establishments that launch their activity in the market have at least 50 employees; this percentage is 7 times higher for establishments that have survived for some time in the market. Large differences in the size of newly created establishments can also be observed across activity sectors.

The principal and most relevant conclusion we present is drawn from the results from the count data models that we have estimated. We have tested, for the purposes of the present research, six types of count data models: Poisson, ZIP, ZIP tau, NB, ZINB, and hurdle model. Each type of model has been run with variables that represent: density of pre-existing establishments in the respective sector, density of large pre-existing establishments, and various municipality characteristics. Care was taken to distinguish outcomes across different activity sectors. We then checked if the results from each model could be improved by incorporating spatial spillovers between municipalities in the form of the distance matrix. In this way, we were able to obtain results from 84 models. In all the analyzed cases, the results from the models run with the distance matrix indicated that accommodating spatial spillovers significantly improves the model's performance. We suggest accounting for spatial spillovers when modeling establishment location choice.

Hurdle models turned out to be preferred. These models offer greater flexibility in modeling zero outcomes and relax the assumption that the zero observations and the positive observations come from the same data generating process. The major difference between hurdle and ZIP models is that the logit component of the hurdle model estimates the probability of a zero count, whereas the logit component of the ZIP

model describes the probability of a zero count from two groups of either always zero and not always zero. Results also indicated that an establishment does not act in isolation and is influenced by other pre-existing establishments. Overall, the greater the presence of establishments from a particular sector, the greater the number of newly created units observed locating nearby in the market. Large establishments tend to attract other industrial units, yet, have a negative effect on other sectors. In addition, when choosing a location in the market, an establishment as a decision-maker may take into consideration not only the characteristics of the potential municipality but also the characteristics of its surroundings due to spatial spillovers from neighboring areas. In addition, we found that new establishments will tend to avoid areas characterized by high real estate prices. Price levels for shops or offices play a strong significant and negative role in location choice for units in construction, commerce, real estate, and for hotels and restaurants. High rates of residence tax appear to discourage the creation of units in the real estate and construction sectors. Establishments will seek areas with high availability of vacant land. The new establishments in the construction sector will tend to concentrate on highly dense residential areas and commercial establishments will try to reach their potential customers. Access to the intellectual workforce is important in the special, scientific and technical sector, in the financial sector as well as for hotels and restaurants. Finally, convenient transport infrastructure seems to play a considerable role in the location choice decisions of establishments.

The contents of the paper may be useful for researchers in the field and those who are interested in policy implications. An approach taken in this paper is aggregate and descriptive and could be further extended in several ways: 1) Other measures beyond simple Euclidean distance (crow-fly distance) may be tested. Free-flow road distance and its corresponding travel time, distance and corresponding travel times at equilibrium during peak hours for private vehicles and public transit may be proposed as alternative weight matrices. 2) The strength of spatial decay of neighborhood effects could be examined. 3) Care should be taken to control for spatial dependence. 4) A challenge would be to develop a disaggregated establishment location choice model incorporating both spatial and strategic interactions between establishments and spatial spillovers among choice alternatives. As discerned by Lambert et al. (2010), accommodation of spatial effects for location decisions may provide a richer, more explicit picture of the regional linkages supporting local growth, industry clustering, and economic development.

Acknowledgements

The authors would like to thank Alice Koegh and Tierra McMahon for their time and editorial support.

References

- [1] Allen, W.B., Liu, D., and S. Singer, 1993. Accessibility measures of US metropolitan areas, *Transportation Research Part B*, 27(6), 439-449.
- [2] Anselin, L., 1988. *Spatial Econometrics: Methods and Models*, Dordrecht, Kluwer Academic Publishers.

- [3] Anselin, L., 2003. Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science Review*, 26, 153-166.
- [4] Arauzo-Carod, J.-M., 2005. Determinants of Industrial Location: An Application for Catalan municipalities, *Papers in Regional Science*, 84(1), 105-120.
- [5] Arauzo-Carod, J.-M., D. Liviano-Solís, and M. Manjón-Antolín, 2010. Empirical studies in industrial location choice: An assessment of their methods and results, *Journal of Regional Science*, 50(3), 685-711.
- [6] Bhat, C.R. and J.Y. Guo, 2004. A Mixed Spatially Correlated Logit Model: Formulation and Application to Residential Choice Modeling, *Transportation Research Part B*, 38(2), 147-168.
- [7] Bodenmann, B.R., 2011. Modelling firm (re-)location choice in UrbanSim, Paper presented at the 51st ERSA Congress, Barcelona, September 2011.
- [8] Bodenmann, B.R. and K.W. Axhausen, 2010. Synthesis report on the state of the art on firmographics, *SustainCity Working Paper 2.3*, Institute for Transport Planning and Systems (IVT), ETH Zurich.
- [9] Bodenmann, B.R. and K.W. Axhausen, 2012. Destination choice of relocating firms - A discrete choice model for the region of St. Gallen, Switzerland, *Papers in Regional Science*, 91(2), 319-341.
- [10] Bondonio, D. and R.T. Greenbaum, 2007. Do local tax incentives affect economic growth? What mean impacts miss in the analysis of enterprise zone policies, *Regional Science and Urban Economics*, 37(1), 121-136.
- [11] Buczkowska, S., N. Coulombel, and M. de Lapparent, 2013. Euclidean distance, network distance, travel time and location choice model of newly created establishments, *Working paper*, Université Paris-Est, Marne la Vallée, France.
- [12] Cameron, A.C. and P.K. Trivedi, 1986. Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests, *Journal of Applied Econometrics*, 1, 29-54.
- [13] Clarke, K.A., 2003. Nonparametric model discrimination in international relations, *Journal of Conflict Resolution*, 47, 72-93.
- [14] Combes, P.P., M. Lafourcade, J.-F. Thisse, and J.-C. Toutain, 2011. The rise and fall of spatial inequalities in France: A long run perspective, *Explorations in Economic History*, 48, 243-271.
- [15] Consul, P.C. and G.C. Jain, 1973. A generalization of Poisson distribution, *Technometrics*, 15, 791-799.
- [16] de Bok, M., 2004. Explaining the location decision of moving firms using their mobility profile and the accessibility of locations, *ERSA conference papers ersa04p338*, European Regional Science Association.
- [17] de Palma, A., 2011. The Grand Paris Project: Tools and Challenges, *International Transport Forum Discussion Papers 2011/28*, OECD Publishing.
- [18] de Palma A., K. Motamedi, J. Moyano, and N. Picard, 2008. Etude des décisions de localisation des emplois en Ile-de-France, *Report for the Direction Régionale de l'Équipement Ile de France (DREIF)*, AdpC, Paris, France.
- [19] Devereux, M.P., R. Griffith, and H. Simpson, 2007. Firm Location Decisions, Regional Grants and Agglomeration Economies, *Journal of Public Economics*, 91, 413-35.

- [20] Garrido, R.A. and H.S. Mahmassani, 2000. Forecasting freight transportation demand with the space-time multinomial probit model, *Transportation Research Part B*, 34(5), 403-418.
- [21] Greene, W.H., 1994. Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models, *Working paper*, Stern School of Business, NYU EC-94-10.
- [22] IAU IdF, 2014. Emploi et crise, Île-de-France et 7 autres régions, January 2014.
- [23] Jayet, H., 2001. Économétrie des données spatiales. Une introduction à la pratique, *Cahiers d'Économie et de Sociologie Rurale*, 58-59, 105-129.
- [24] Klier, T. and D. McMillen, 2008. Clustering of auto supplier plants in the United States, *Journal of Business & Economic Statistics*, American Statistical Association, 26(4), 460-471.
- [25] Lambert, D., 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing, *Technometrics*, 34, 1-14.
- [26] Lambert, D.M., J.P. Brown, and R.J.G.M. Florax, 2010. A two-step estimator for a spatial lag model of counts: theory, small sample performance and an application, *Regional Science and Urban Economics*, 40(4), 241-252.
- [27] Liviano-Solís, D. and J.-M. Arauzo-Carod, 2011. Industrial Location and Space: New Insights, *Working Papers 2072/152137*, Universitat Rovira i Virgili, Department of Economics.
- [28] Maoh, H.F., 2005. *Modeling firm demography in urban areas with an application to Hamilton, Ontario: towards an agent-based approach*, PhD Thesis, School of Geography and Earth Sciences, McMaster University, Hamilton, Canada.
- [29] Maoh, H.F. and P.S. Kanaroglou, 2005. Agent-Based Firmographic Models: a Simulation Framework for the City of Hamilton, Proceedings of PROCESSUS Second International Colloquium on the Behavioural Foundations of Integrated Land-use and Transportation Models: Frameworks, Models and Applications, Toronto.
- [30] Maoh, H.F. and P.S. Kanaroglou, 2007. Business establishment mobility behavior in urban areas: a microanalytical model for the City of Hamilton in Ontario, Canada, *Journal of Geographical Systems*, 9(3), 229-252.
- [31] Miyamoto, K., V. Vichiensan, N. Shimomura, and A. Paez, 2004. Discrete choice model with structuralized spatial effects for location analysis. In: *Transportation Research Record: Journal of the Transportation Research Board*, 312, TRB, National Research Council, Washington, DC.
- [32] Mohammadian, A.K., M. Haider, and P.S. Kanaroglou, 2005. Incorporating spatial dependencies in random parameter discrete choice models, in *TRB* (ed.) the 84th Annual Meeting of the Transportation Research Board, Washington, D.C.
- [33] Motamedi, K., 2008. *Contribution à la modélisation intégrée de la dynamique régionale et des transports: application à la région Parisienne*, PhD dissertation, University of Cergy-Pontoise, Cergy, France.
- [34] Mullahy, J., 1986. Specification and testing of some modified count models. *Journal of Econometrics*, 33, 341-365.

- [35] Neumark, D. and J. Kolko, 2010. Do enterprise zones create jobs? Evidence from California's enterprise zone program, *Journal of Urban Economics*, 68(1), 1-19.
- [36] Nguyen, C.Y., K. Sano, T.V. Tran, and T.T. Doan, 2012. Firm relocation patterns incorporating spatial interactions, *The Annals of Regional Science*, Published online: 11 August 2012.
- [37] Rocha, N., 2008. Firm Location Determinants: Empirical Evidence for France, *IHEID Working Papers 08-2008*, Economics Section, The Graduate Institute of International Studies.
- [38] Sener, I., R. Pendyala, and C. Bhat, 2011. Accommodating spatial correlation across choice alternatives in discrete choice models: an application to modeling residential location choice behavior, *Journal of Transport Geography*, 19(2), 293-303.
- [39] Smirnov, O., 2010. Modeling Spatial Discrete Choice, *Regional Science and Urban Economics*, 40(5), 292-298.
- [40] Siebert, H., 2000. The paradigm of locational competition, *Kieler Diskussionsbeiträge*, 367, Institut für Weltwirtschaft, Kiel.
- [41] Strotmann, H., 2007. Entrepreneurial Survival, *Small Business Economics*, 28, 87-104.
- [42] Sutton, J., 2007. Market Structure: Theory and Evidence, Elsevier, vol. 3 of *Handbook of Industrial Organization*, chap. 35, 2301-2368.
- [43] van Wissen, L.J., 2000. A micro-simulation model of firms: Applications of concepts of the demography of the firm, *Papers in Regional Science*, 79, 111-134.
- [44] Vichiensan, V., Y. Tokunaga, and K. Miyamoto, 2005. Mixed logit model framework with structuralized spatial effects: a test of applicability with area unit systems in location analysis, *Journal of the Eastern Asia Society for Transportation Studies*, 6, 3789-3802
- [45] Winkelmann, R., 2003. *Econometric Analysis of Count Data*, 4th ed. Springer Verlag, Heidelberg.

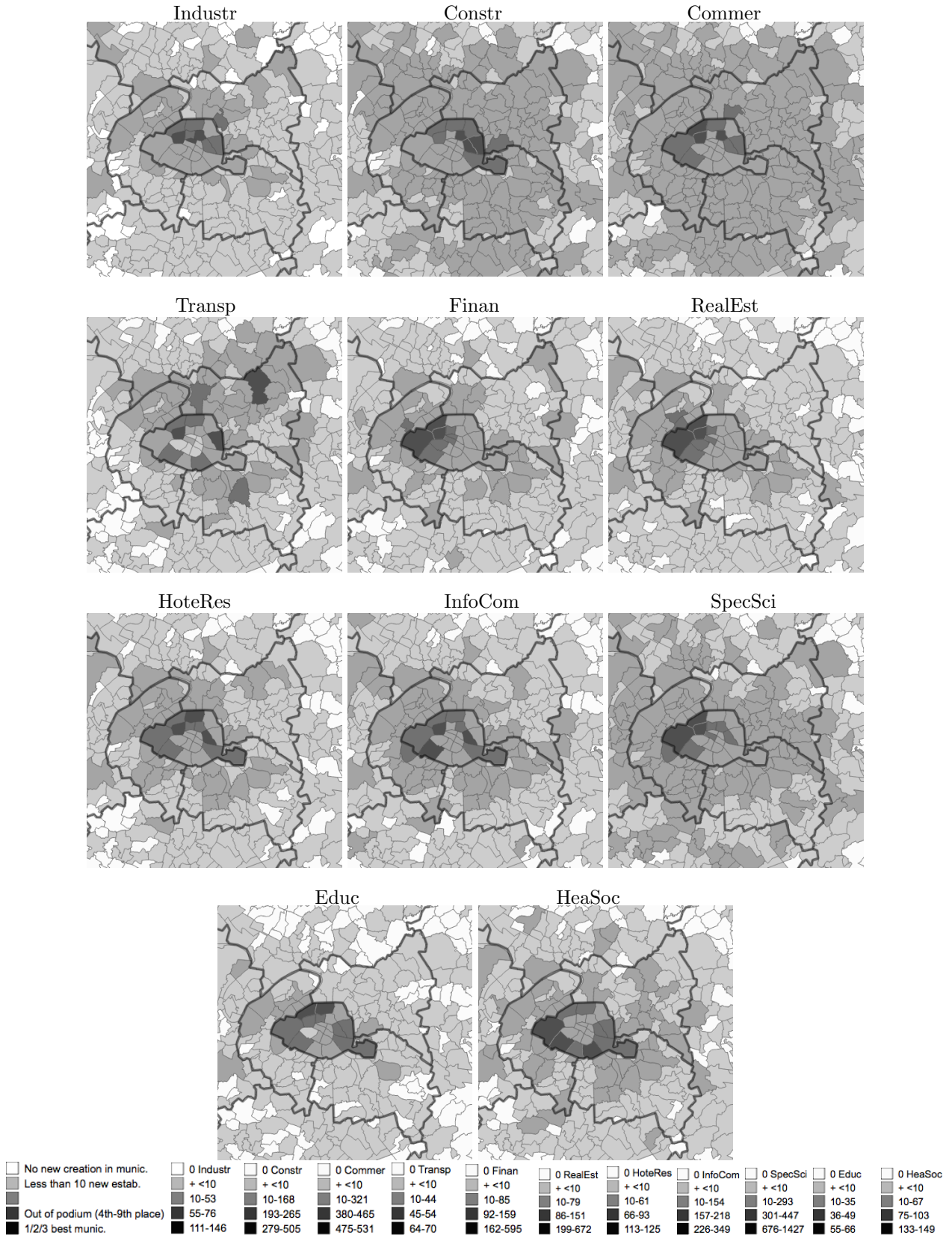


Figure 1: Newly created establishments in 2007 - rankings based on the market share - focus on Paris and the inner ring

Table 4: Hurdle model with distance matrix: explanatory variables represent density of pre-existing establishments in the respective sector, density of large pre-existing establishments, and various municipality characteristics

Hurdle part I: zero counts	Industr ^a		Constr		Commer		Finan		RealEst		HoteRes		SpecSci	
	Est	T-st ^b	Est	T-st	Est	T-st	Est	T-st	Est	T-st	Est	T-st	Est	T-st
Constant	-18,330***	-6,51	-23,338***	-7,20	-23,420***	-6,76	-16,307***	-5,32	-9,949**	-2,06	-24,271***	-6,72	-20,789***	-5,37
Estab. from respective sector ^c	-1,218***	-12,29	-0,663***	-5,31	-1,053***	-10,51	-0,580***	-6,27	-0,931***	-9,69	-1,374***	-12,59	-0,882***	-8,66
Large estab. from all sectors ^d	0,053	1,10	0,049	1,16	0,003	0,06	-0,187**	-2,42	0,108*	1,83	0,066	1,15	0,046	0,86
White-collar workers							-0,283**	-2,34			-0,155	-1,45	-0,205**	-2,18
Blue-collar workers	0,070	0,62												
Trips home-work					-0,043	-0,20								
Population density			-1,043***	-4,13					0,123	0,29			-0,180	-0,79
Income per person														
Offices														
Shops and offices														
Vacant land	0,136	1,50	0,096	1,08	-0,032	-0,35	0,010	0,16	-0,098*	-1,76				
Distance to highway	0,114	1,28	0,117	1,36	0,006	0,06	0,070	0,70			0,311***	2,93		
Public transport			-0,052*	-1,84	-0,021	-0,73	-0,042	-0,41			-0,005	-0,05	0,057	0,62
Residence tax			0,005	0,02			-0,092**	-2,39			-0,010	-0,28	-0,091***	-2,88
Price of offices	2,418***	6,65	3,016***	7,49			1,937***	5,10					2,485***	5,22
Price of shops					2,814***	6,81					3,102***	7,09		

^aDependent variable: number of newly created establishments in 2007.

^b***, **, * represent statistical significance at the 1%, 5%, and 10% level, respectively.

^cThe data on stock of establishments used to construct the explanatory variables in the model are given for the

1st of January 2007. See Table 1 for description of potential explanatory variables.

^dLarge pre-existing establishments from all sectors with 50 employees or more.

Table 5: Hurdle model with distance matrix: explanatory variables represent density of pre-existing establishments in the respective sector, density of large pre-existing establishments, and various municipality characteristics (continuation)

	Industr ^a		Constr		Commer		Finan		RealEst		HoteRes		SpecSci	
	Est	T-st ^b	Est	T-st	Est	T-st	Est	T-st	Est	T-st	Est	T-st	Est	T-st
Constant	0,269	0,44	12,596***	38,94	13,470***	42,49	3,255***	4,01	11,621***	15,78	5,857***	7,03	2,720***	6,94
Estab. from respective sector ^c	0,593***	15,11	1,148***	54,85	0,527***	31,21	1,084***	20,21	0,812***	18,20	0,600***	12,10	1,001***	40,45
Large estab. from all sectors ^d	0,087**	2,49	-0,128***	-9,37	-0,060***	-4,69	-0,760***	-16,18	-0,203***	-4,76	-0,228***	-4,89	-0,131***	-6,16
White-collar workers							0,771***	19,17			0,307***	6,14	0,154***	5,70
Blue-collar workers	0,083**	2,08												
Trips home-work			0,134***	4,09	1,130***	35,10								
Population density														
Income per person														
Offices									0,377***	6,53			-0,865***	-29,34
Shops and offices														
Vacant land	0,064**	2,30	0,072***	5,01	0,049***	4,43	-0,050	-1,31	0,240***	5,91	0,146***	5,11		
Distance to highway	-0,215***	-6,19	-0,010	-0,58	-0,311***	-24,67	-0,478***	-16,18						
Public transport			0,225***	18,84	0,260***	27,82	0,313***	12,25	0,251***	10,32	0,501***	16,47	-0,562***	-38,46
Residence tax			-1,312***	-30,47					-2,137***	-32,30			0,430***	31,39
Price of offices	0,013	0,16	-1,154***	-29,53	-1,220***	-31,81	-0,188*	-1,80	-0,610***	-7,29	-0,576***	-5,41	-0,072	-1,38
Price of shops														
Log-Lik	-2071,0		-4472,4		-5194,5		-2214,4		-2449,1		-1803,5		-4424,4	
AIC	4169,9		8980,8		10421,0		4464,8		4930,2		3638,9		8884,7	
AICC	4170,3		8981,3		10422,0		4465,3		4930,7		3639,3		8885,3	
BIC	4242,3		9073,9		10504,0		4557,9		5013,0		3721,6		8977,8	
Convergence	Yes		Yes		Yes		Yes		Yes		Yes		Yes	
#Parameters	2x7		2x9		2x8		2x9		2x8		2x8		2x9	

^aDependent variable: number of newly created establishments in 2007.

^b***, **, * represent statistical significance at the 1%, 5%, and 10% level, respectively.

^cThe data on stock of establishments used to construct the explanatory variables in the model are given for the 1st of January 2007. See Table 1 for description of potential explanatory variables.

^dLarge pre-existing establishments from all sectors with 50 employees or more.